LEVEL II

UNCLASS

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFIT-CI-79-244D	2. GOVT ACCESSION NO.: AD-A091008
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): Linear Multistep Methods with Near-Optimal Stability.	5. TYPE OF REPORT & PERIOD COVERED: THESIS/DISSERTATION
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): David Jerome Nolting	8. CONTRACT OR GRANT NUMBER(s): Doctoral thesis
9. PERFORMING ORGANIZATION NAME AND ADDRESS: AFIT STUDENT AT: Univ of Missouri-Columbia	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: AFIT/NR, WPAFB OH 45433	12. REPORT DATE: Dec 79
	13. NUMBER OF PAGES: 97
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 12 198	15. SECURITY CLASS. (of this report): UNCLASS
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

DTIC ELECTE OCT 29 1980 D

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

APPROVED FOR PUBLIC RELEASE: IAW AFR 190-17

25 SEP 1980

Fredric C. Lynch

FREDRIC C. LYNCH, Major, USAF
Director of Public Affairs
Air Force Institute of Technology (ATC)
Wright-Patterson AFB, OH 45433

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

ATTACHED

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASS

012200

AD A091008

DDC FILE COPY

80 10 14 234

LINEAR MULTISTEP METHODS WITH NEAR-OPTIMAL STABILITY

David Jerome Nolting

Prof. David J. Rodabaugh Dissertation Supervisor

ABSTRACT

Desirable properties of linear multistep methods (LMM) can be optimized by viewing those properties as functional values and the LMM possessing those properties as points in a domain space. This study conducts two such optimizations numerically. The first is a search for relatively stable explicit LMM and the second is a search for stiffly stable implicit LMM. Near-optimally relatively stable explicit LMM are found for orders four through nine.

In the second study, the concept of $A(\alpha, r)$-stability is introduced for stiffly stable LMM. It recognizes the need for large regions of absolute stability in the left half plane and the need for a region of accuracy about the origin defined by the region of relative stability. An economical means of determining the region of relative stability is developed and used. Nearly-optimal $A(\alpha, r)$-stable implicit LMM are found for orders four through six for a variety of classes determined by fixed error constants $C_{p+1}/\sigma(1)$.

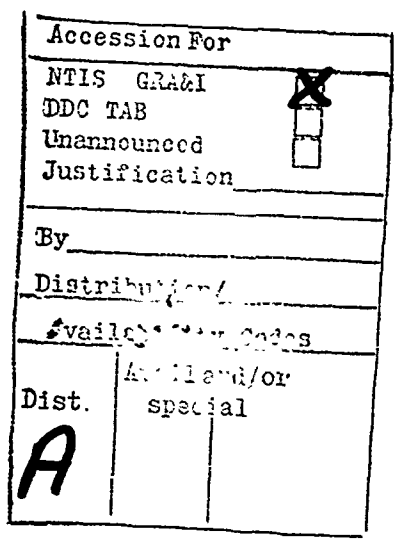
LINEAR MULTISTEP METHODS WITH NEAR-OPTIMAL STABILITY

A Dissertation

Presented to

the Faculty of the Graduate School

University of Missouri-Columbia



In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

David Jerome Nolting

December, 1979

Prof. David J. Rodabaugh Dissertation Supervisor

TABLE OF CONTENTS

CHAPTER I

PRELIMINARIES

1.1 Introduction

Consider the initial value problem

$$y' = f(x, y),\ x\varepsilon[a, b],\ y(a) = y_0 \qquad (1.1\text{-}1)$$

where y, y_0, and $f(x, y)$ are in R^n and sufficient conditions are placed on f to ensure a unique solution exists. We seek numerical methods to solve the problem above more accurately and to solve larger classes of such problems. Especially when $n > 1$ in the problem given, it can be roughly classified as stiff or non-stiff. We investigate both classes and find improved methods for solving members of each class.

Two basic approaches to a numerical approximation of the solution to (1.1-1) are the linear multistep and Runge-Kutta methods. Most, if not all, popular approaches have stemmed from one or both of these methods. We limit the ensuing investigation to linear multistep methods (LMM).

1.2 Linear Multistep Approach

When it is not possible to solve the continuous system (1.1-1) exactly we must be content with a discrete model of the given system. We restrict ourselves to an evenly spaced mesh of

the interval [a, b]. Denote this mesh by $\{x_0 = a, \cdots,$ $x_i = a + ih, \cdots, x_n = b\}$ so that we have constant mesh-width

$$h = \frac{b - a}{N} .$$

Define $y_n = y(x_n)$ and $y'_n = f(x_n, y_n)$, and denote our respective approximations by y_n and $y'_n = f(x_n, y_n)$. Our discrete model then consists of $\{y_i | i = 0, \cdots, N\}$.

A linear k-step method expresses a linear relationship between y_{n+1}, y'_{n+1} and the previous k values of y_i and y'_i. This linear relationship takes the form

$$y_{n+1} = \sum_{i=0}^{k-1} \alpha_i y_{n-i} + h \sum_{i=-1}^{k-1} \beta_i y'_{n-i} \qquad (1.2\text{-}1)$$

for real values of α_i and β_i, where $|\alpha_{k-1}| + |\beta_{k-1}| \neq 0$. Returning to the continuous model, consider

$$\sum_{i=-1}^{k-1} \alpha_i y_{n-i} + \sum_{i=-1}^{k-1} \beta_i y'_{n-i},$$

where $\alpha_{-1} = 1$. If $y(x)$ is sufficiently differentiable, by a Taylor expansion about x_n for y_{n-i} and y'_{n-i}, $i = -1, 0, 1, \cdots, k-1$, we get

$$\sum_{i=-1}^{k-1} (\alpha_i y_{n-i} + h\beta_i y'_{n-i}) = \sum_{i=0}^{\infty} C_i h^i y^{(i)}(x_n) \qquad (1.2\text{-}2)$$

where the constants C_i are determined from the coefficients α_i and β_i of the LMM (1.2-1). We have the following expressions for the C_i:

$$C_0 = \sum_{i=-1}^{k-1} \alpha_i,$$

$$C_1 = \sum_{i=-1}^{k-1} (-i)\alpha_i + \sum_{i=-1}^{k-1} \beta_i, \tag{1.2-3}$$

$$C_p = \frac{1}{p!} \sum_{i=-1}^{k-1} (-i)^p \alpha_i + \frac{1}{(p-1)!} \sum_{i=-1}^{k-1} (-i)^{p-1} \beta_i \text{ for } p = 2, 3, \cdots .$$

The <u>order</u> of a LMM (1.2-1) is defined to be the smallest integer p such that $C_i = 0$ for $0 \leq i \leq p$ and $C_{p+1} \neq 0$. When the order is at least one the LMM is said to be <u>consistent</u>. The following concept of convergence is critical for any useful LMM.

<u>Definition</u>: A LMM is said to be <u>convergent</u> if for any problem (1.1-1) for which we are guaranteed a unique solution, we have

$$\lim_{\substack{n\to 0 \\ nh = x-a}} y_n = y(x)$$

for all $x\varepsilon[a, b]$ and for all solutions $\{y_n\}$ of (1.2-1) satisfying starting conditions $y_i = \eta_i(h)$ where

$$\lim_{h\to 0} \eta_i(h) = y_0, \; i = 0, 1, \cdots , k-1.$$

With the LMM (1.2-1) we associate two polynomials $\rho(z)$ and $\sigma(z)$. Define

$$\rho(z) = \sum_{i=-1}^{k-1} \alpha_i z^{k-(i+1)}$$

and (1.2-4)

$$\sigma(z) = \sum_{i=-1}^{k-1} \beta_i z^{k-(i+1)},$$

where $\alpha_{-1} = -1$. Then any LMM (1.2-1) uniquely determines both ρ and σ and conversely. If $\rho(z)$ has no roots greater than one in modulus and if all roots of modulus one are simple, the associated LMM is said to be zero-stable. We have the following theorem (stated without proof) from Dahlquist [5].

Theorem 1.2-1. A LMM is convergent if and only if it is consistent and zero-stable.

Formula (1.2-1) provides an explicit means of obtaining y_{n+1} when $\beta_{-1} = 0$. The resulting LMM are called predictors. When $\beta_{-1} \neq 0$ the formula defines y_{n+1} implicitly. In this case iterative values

$$\{y_{n+1}^{[s]}\}_{s=1}^{\infty}$$

are computed via the relationship

$$y_{n+1}^{[s+1]} = \phi(y_{n+1}^{[s]}), \quad (1.2\text{-}5)$$

where ϕ is defined by the right-hand side of (1.2-1). As long as ϕ is contractive the sequence

$$\{y_{n+1}^{[s]}\}_{s=1}^{\infty}$$

will converge to y_{n+1}. This iteration, normally used to obtain the value of y_{n+1} when $\beta_{-1} \neq 0$, prompts the name correctors for this class of LMM. For any k-step method k starting values are required to implement the recursive calculation of the y_i, $i \geq k$. These values are normally found by some explicit LMM or R-k method.

Denote the class of n^{th} order, m step predictors and correctors respectively by P(n, m) and C(n, m). For example the n^{th} order Adams-Moulton formula is a member of C(n, n-1). Similarly the n^{th} order Adams-Bashforth formula is a member of p(n, n). We search within p(n, n), n = 4, ··· , 9 for improved methods on non-stiff problems and within C(n, n), n = 4, 5, 6 for improved methods on stiff problems.

1.3 Coefficient Matrix Derivation

Let *L* be a LMM of the class C(n, n). As in (1.2-1) we normally assume $\alpha_{-1} = 1$ to eliminate ambiguity between equivalent LMM. Then from (1.2-3) we have n+1 independent equations relating 2n + 1 parameters. This leaves n unspecified (free) parameters to determine *L*. We take these free parameters to be $\alpha_1, \cdots, \alpha_{n-1}$, and β_{-1}. In (1.2-3) $C_0 = 0$ determines α_0 and $C_1 = \cdots = C_n = 0$ yields n x n matrices B and D and an n x (n+1) matrix A such that $DA[1, \alpha_1, \cdots, \alpha_{n-1}, \beta_{-1}]^T = B[\beta_0, \cdots, \beta_{n-1}]^T$. These matrices have the form

$$A = \begin{bmatrix} 1 & 1 & 2 & \cdots & (n-1) & -1 \\ -1 & 1 & 2^2 & \cdots & (n-1)^2 & 2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ (-1)^{n-1} & 1 & 2^n & \cdots & (n-1)^n & (-1)^n n \end{bmatrix}, \quad (1.3\text{-}1)$$

$$B = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 2 & \cdots & (n-1) \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & 2^{n-1} & \cdots & (n-1)^{n-1} \end{bmatrix},$$

and $D = (d_{ij})$ is diagonal with $d_{ii} = 1/i$. In this section we define B^{-1} recursively, that is, we define the B^{-1} for $C(n, n)$ from information concerning the B^{-1} for $C(n-1, n-1)$. Thereby we have the means to calcuate the exact rational entries of the $n \times (n+1)$ matrix $B^{-1}DA$ and so all coefficients of L are explicitly determined by the n free parameters.

Define the $n \times n$ matrix S^n recursively as follows. Let

$$S^1 = (1), \; S^2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

and $S^n = (\sigma^n_{ij})$ where

$$\sigma^n_{1,\,1} = (n-1)! \; , \; \sigma^n_{i,\,1} = 0 \text{ for } 2 \leq i \leq n,$$

$$\sigma^n_{i,\,n} = 1 \text{ for } 1 \leq i \leq n, \; \sigma^n_{n,\,j} = \sigma^{n-1}_{1,\,j-1} \text{ for } 2 \leq j \leq n-1,$$

and $\sigma^{n}_{i,j} = \sigma^{n-1}_{i,j-1} + (n-1)\,\sigma^{n-1}_{i,j}$ for $1 \leq i \leq n-1$ and $2 \leq j \leq n-1$.

Denote the B^{-1} for $C(n, n)$ by $B_n^{-1} = (b^n_{ij})$. Then we have

$$b^n_{i,j} = \frac{(-1)^{i+j}\,\sigma^r_{i,j}}{(i-1)!(n-i)!} \quad \text{for } 2 \leq i, j \leq n,$$

$$b^n_{1,1} = 1,\ b^n_{1,j} = -\sum_{\ell=2}^{n} b^n_{\ell,j} \quad \text{for } 2 \leq j \leq n,$$

and

$$b^n_{i,1} = 0 \quad \text{for } 2 \leq i \leq n.$$

When computing the exact rational entries of the matrix $B_n^{-1}DA$ for higher orders it is necessary to increase the machine precision and use careful programming. We have calculated rational entries for these matrices of coefficients through order 18. In Appendix A we list those matrices for orders two through nine.

CHAPTER II

STABILITY ANALYSIS

2.1 A General Consideration

Many questions concerning the stability of LHM reduce to determining whether certain polynomials have roots less than one in modulus. Much effort towards this end has been devoted to transformed polynomials whose roots lie in the open left half plane. We say a polynomial is of the Schur type if it has roots less than one in modulus and of the Hurwitz type if it has all roots lying in the open left half plane. Several forms of necessary and sufficient conditions exist in the literature for a polynomial to be Hurwitzian [23, 26]. Translation to necessary and sufficient conditions for Schur polynomials normally leads to intractable criteria. We manage to find one exception to this, although in the general case it provides sufficiency only. First we establish several preliminary results.

Lemma 2.1-1. Let A be an $(n+1) \times n$ matrix and $c \in R^n$. Also let the set of points $x \in R^n$ satisfying the system of linear inequalities $Ax \geq C$ define an n-dimensional simplex Σ_n with non-empty interior. Let τ be the set of all points $y \in R^n$ such that $Ay < C$. Then τ is empty.

Proof: Let x_0 be a point interior to Σ_n and suppose there exists a point $y_0 \in \tau$. Consider the line segment $\alpha x_0 + (1-\alpha) y_0$ where

$0 \leq \alpha \leq 1$. This line segment must intersect each of the n+1 hyperplanes defining Σ_n since $Ax_0 > C$ and $Ay_0 < C$.

Now extend this line segment to $\alpha > 1$. Since Σ_n is bounded this extension must intersect at least one of the defining hyperplanes a second time at a point distinct from the first point of intersection. Thus we have the line $\alpha x_0 + (1-\alpha)y_0$ contained in one of the defining hyperplanes of Σ_n. This contradicts the assumptions that $Ax_0 > C$ and $Ay_0 < C$. QED.

Lemma 2.1-2. Let

$$P(z) = \sum_{i=0}^{n} \beta_i z^i$$

be a Hurwitz polynomial. Then all β_i, $i = 0, 1, \cdots, n$ have the same sign.

Proof: If all roots of P have negative real part, then

$$P(z) = \beta_n \prod_{i=1}^{k} (z + r_i) \prod_{i=1}^{\frac{(n-k)}{2}} ((z + C_i)^2 + d_i^2)$$

where $r_i > 0$ and $C_i > 0$ so that all coefficients β_i take the sign of β_n. QED.

Lemma 2.1-3. Consider the mapping

$$z \to \frac{z + 1}{z - 1}$$

which maps the unit disk onto the left half plane. For any polynomial

$$p(z) = \sum_{i=0}^{n} a_i z^i$$

define a new polynomial

$$P(z) = (z-1)^n \, p\left(\frac{z+1}{z-1}\right).$$

Then for the polynomial transformation $T : p \to P$ we have $T^2 = 2^n I$ where I is the identity transformation.

Proof:

$$T^2(p) = T(P) = (z-1)^n \, P\left(\frac{z+1}{z-1}\right) = (z-1)^n \left(\frac{2}{z-1}\right)^n p\left(\frac{2z}{2}\right)$$

$$= 2^n \, p(z) = 2^n \, I(p). \quad \text{QED.}$$

Consider the real monic polynomial

$$p(z) = \sum_{i=0}^{n} \alpha_i z^i, \; \alpha_n = 1.$$

Let the coefficient vector of p be $\bar{\alpha} = (\alpha_0, \cdots, \alpha_{n-1})$ and $\bar{\alpha}_j$ be the coefficient vector of $(z-1)^{n-j} \, (z+1)^j$. Define S_n to be the n-dimensional simplex determined by $\bar{\alpha}_j$, $j = 0, 1, \cdots, n$.

Theorem 2.1-4. The coefficient vector of any n^{th} order real monic Schur polynomial lies in S_n.

Proof: Let

$$p(z) = \sum_{i=0}^{n} a_i z^i, \; a_n = 1,$$

be a real monic Schur polynomial. Transform the roots of p to the left half plane via the mapping

$$z \to \frac{z+1}{z-1}.$$

Define the new polynomial

$$P(z) = \sum_{i=0}^{n} \beta_i z^i$$

as in Lemma 2.1-2. Then

$$P(z) = \sum_{i=0}^{n} a_i (z-1)^{n-i}(z+1)^i$$

and

$$\beta_i = \sum_{j=0}^{n} a_j \sum_{\substack{k=0 \\ 0 \le i-k \le j}}^{n-j} \binom{n-j}{k} \binom{j}{i-k} (-1)^{n-j-k}, \quad i = 0, 1, \cdots, n.$$

From this expression denote the coefficient of z^i in $(z-1)^{n-j}(z+1)^j$ by M^n_{ij} and let $M_n = (m^n_{ij})$, $0 \le i, j \le n$. We have then a matrix equation $M_n(a_0, \cdots, a_n)^T = (\beta_0, \cdots, \beta_n)^T$. We note that M_n can also be described by

$$m^n_{i0} = (-1)^{n-1} \binom{n}{i}, \quad i = 0, 1, \cdots, n$$

$$m^n_{nj} = 1, \quad j = 0, 1, \cdots, n$$

and

$$m^n_{ij} = m^n_{i,\,j-1} + m^n_{i+1,\,j-1} + m^n_{i+1,\,j}, \quad \begin{matrix} i = 0, 1, \cdots, n-1 \\ j = 1, 2, \cdots, n. \end{matrix}$$

Lemma 2.1-3 provides $M_n^2 = 2^n I_{n+1}$, where I_{n+1} is the (n+1) x (n+1) identity matrix.

Since p(z) is a real monic polynomial, the first n elements in the j^{th} column of M_n is the vector $\bar{\alpha}_j$ described in the discussion previous to the statement of the theorem. Thus taking the product M_n^2 corresponds to evaluating n+1 hyperplanes in R^n at the coefficient vectors $\bar{\alpha}_j$, $j = 0, 1, \cdots, n$. Since $M_n^2 = 2^n I_{n+1}$ each of the coefficient vectors $\bar{\alpha}_j$ lies on all but one of the n+1 hyperplanes. They therefore describe S_n.

Let f_k represent the k^{th} hyperplane where for $x \in R^n$ the k^{th} row of $M_n(x_1, \cdots, x_n, 1)^T$ defines $f_k(x)$. We define the positive side of the k^{th} hyperplane to consist of those points $x \in R^n$ for which $f_k(x) > 0$. $M_n^2 = 2^n I_{n+1}$ implies each $\bar{\alpha}_j$ lies on the positive side of its opposite face. Thus for each $x \in R^n$ we have $x \in S_n$ if and only if $f_k(x)$ is positive for each $k = 1, \cdots, n + 1$.

Notice if $\bar{0}$ is the zero of R^n we have $f_k(\bar{0}) > 0$ for each $k = 1, \cdots, n + 1$, so that continuity of the f_k provide S_n with a non-empty interior.

Since p is a Schur polynomial, P(z) is a Hurwitz polynomial and lemma 2.1-2 provides that all β_i are of the same sign. Lemma 2.1-1 excludes the possibility that all β_i are negative. Since $\beta_i = (a_0, a_1, \cdots, a_{n-1})$ we have the coefficient vector of p(z) in S_n. QED.

This Theorem is in fact necessary and sufficient when $n \leq 2$. We obtain the following useful corollary.

Corollary 2.1-5. Let

$$p(z) = \sum_{i=0}^{n} a_i z^i$$

and N_n be as in the proof of Theorem 2.1-4. Let

$$P(z) = \sum_{i=0}^{n} \beta_i z^i$$

where $N_n(a_0, a_1, \cdots, a_n)^T = (\beta_0, \beta_1, \cdots, \beta_n)^T$. Then if $p(z)$ is a Schur polynomial all the β_i have the same sign

Proof: From the proof of the Theorem we see that $P(z)$ is a Hurwitz polynomial. The result follows upon application of Lemma 2.1-2. QED.

2.2 Stability for Non-Stiff LMM

In Chapter I we mentioned convergence of a LMM and its equivalence to consistency and zero-stability. There we were concerned about what happens to the error through successive calculations at a fixed point within the interval of integration as $h \to 0$ and $n \to \infty$. The definition of convergence requires this error to go to zero on all problems of a certain class given sufficiently accurate starting values. Now we concern ourselves with what happens to the error through successive calculations using a fixed step length h as we proceed through the interval of integration. Odeh and Liniger [30] refer to this concept as fixed-h stability.

Definition: A LMM is called fixed-h stable if the accumulated truncation error in solving the model equation $y' = \lambda y$,

where λ is a complex constant, remains bounded as $n \to \infty$ for a non-vanishing range of values of the constant $q = \lambda h$.

Here we come to a disjunction in the stability analyses for stiff and non-stiff LMM. Non-stiff methods normally have bounded regions of fixed-h stability in the q-plane and stiff methods require those regions to be unbounded. We pursue the analysis for stiff methods more fully in the next section.

The concepts of absolute stability and relative stability are more commonly used. We denote the characteristic polynomial of a LMM by $\Pi(z, q) = \rho(z) + q\sigma(z)$ where ρ and σ are taken from (1.2-4) and q is defined as in the previous definition from the model test equation $y' = \lambda y$.

Definition: A LMM is absolutely stable at a given point q_0 if all roots of $\Pi(z, q_0)$ lie inside the unit circle. The region of absolute stability for a LMM is the set of all points q_0 in the q-plane at which the method is absolutely stable.

Absolute stability forces solution of the model test equation to have a decreasing global error and is therefore more restrictive than fixed-h stability.

For any consistent and zero-stable method L, $\Pi(z, 0) = \rho(z)$ has a simple root at $z = 1$. This root is called the principal root and is denoted by r_1. As long as the leading coefficient of a polynomial is non-zero, its roots are continuous functions of the coefficients. Thus we may follow this principal root r_1 to non-zero values of q. For q sufficiently near zero we have

$$r_1 = e^q + O(q^{k+1}) \qquad (2.1\text{-}1)$$

whenever L is of order k. For an argument establishing this relationship see Lambert [27, p. 66].

Definition: A LMM is relatively stable at a given point q_0 if the root of $\Pi(z, q_0)$ of largest magnitude is the principal root. The region of relative stability for a LMM is the set of all points q_0 in the q-plane for which the method is relatively stable.

There are numerous definitions of relative stability given in the literature (for example see [27, p. 68]); in light of (2.1-1) the definition above is very useful. We take a closer look at alternate definitions in Chapter IV. Relative stability limits the rate of growth of the global error to approximately that of the solution. We will examine this relative error more closely in Chapter III.

For methods which are consistent and zero-stable there will not be regions of absolute stability near the origin in the right half plane as there are with relative stability, while in the left half plane the regions of absolute stability are larger than those of relative stability. These conclusions follow since for q arbitrarily near the origin in the right half plane we have $r_1 > 1$, and for q in the left half plane $r_1 < 1$. When solving a problem with q lying in the left half plane and inside the region of absolute stability but outside the region of relative stability, the global error goes to zero but possibly much slower than the solution

itself. We illustrate this possibility with an example in Chapter IV. Our results must then be viewed with this is mind.

Throughout the remainder of the section we assume $\rho(z)$ and $\sigma(z)$ have no roots in common and that $\sigma(z)$ has no roots of unit magnitude. To find the boundary of the region of absolute stability we take the image of the unit circle under the map

$$z \to q = -\frac{\rho(z)}{\sigma(z)} .$$

This boundary is not always a simple closed curve. The boundary of the region of relative stability is found numerically by tracking the principal root on various rays emanating from the origin. These notions are considered more carefully in Chapter IV. The following result is often used, many times casually, and we now state it formally.

Lemma 2.2-1. The boundary of the regions of absolute and relative stability are symmetric about the real axis.

Proof: If q and $\bar{q}$ are complex conjugates we have $\overline{\Pi(z, q)} = \Pi(\bar{z}, \bar{q})$ so that z_0 is a root of $\Pi(z, q)$ if and only if $\bar{z}_0$ is a root of $\Pi(z, \bar{q})$. Thus the sets of moduli of the roots of $\Pi(z, q)$ and $\Pi(z, \bar{q})$ correspond and also $|z_1(q)| = |z_1(\bar{q})|$ where $z_1(q)$ is the principal root of $\Pi(z, q)$. Since the stability boundaries are based on the moduli of the roots of $\Pi(z, q)$ the result follows. QED.

We now take a closer look at the boundary of the region of absolute stability and develop several results. We use the following notation.

$$\rho_r(\theta) = \mathrm{Re}(\rho(e^{i\theta})) = \sum_{j=-1}^{k-1} \alpha_j \cos\,(k-(j+1))\theta$$

$$\rho_i(\theta) = \mathrm{Im}(\rho(e^{i\theta})) = \sum_{j=-1}^{k-1} \alpha_j \sin\,(k-(j+1))\theta$$

$$\sigma_r(\theta) = \mathrm{Re}(\sigma(e^{i\theta})) = \sum_{j=-1}^{k-1} \beta_j \cos\,(k-(j+1))\theta$$

and

$$\sigma_i(\theta) = \mathrm{Im}(\sigma(e^{i\theta})) = \sum_{j=-1}^{k-1} \beta_j \cos\,(k-(j+1))\theta.$$

Then $\rho(e^{i\theta}) = \rho_r(\theta) + i\rho_i(\theta)$ and $\sigma(e^{i\theta}) = \sigma_r(\theta) + i\sigma_i(\theta)$. We assume no ambiguity exists between the subscript i and the complex number i. Let α_θ be the angle measured from the negative real axis clockwise to the line through

$$-\frac{\rho(e^{i\theta})}{\sigma(e^{i\theta})}$$

and the origin, for $0 \leq \theta \leq \Pi$. Since

$$\mathrm{Re}\left[-\frac{\rho(e^{i\theta})}{\sigma(e^{i\theta})}\right] = -\frac{\rho_r(\theta)\sigma_r(\theta) + \rho_i(\theta)\sigma_i(\theta)}{\sigma_r^{\,2}(\theta) + \sigma_i^{\,2}(\theta)}$$

and

$$\text{Im}\left[-\frac{\rho(e^{i\theta})}{\sigma(e^{i\theta})}\right] = -\frac{\rho_i(\theta)\sigma_r(\theta) - \rho_r(\theta)\sigma_i(\theta)}{\sigma_r^2(\theta) + \sigma_i^2(\theta)},$$

we can express α_θ as

$$\alpha_\theta = \cot^{-1}\left[\frac{\rho_r(\theta)\sigma_r(\theta) + \rho_i(\theta)\sigma_i(\theta)}{\rho_r(\theta)\sigma_i(\theta) - \sigma_r(\theta)\rho_i(\theta)}\right]. \qquad (2.1\text{-}2)$$

Lemma 2.2-2. For any convergent LMM there is a deleted neighborhood of zero in which $\rho_i(\theta) \neq 0$.

Proof: We have

$$\rho_i'(\theta) = \sum_{j=-1}^{k-1} (k-j-1)\alpha_j \cos(k-j-1)\theta.$$

Now

$$\rho_i'(0) = \sum_{j=-1}^{k-1} (k-j-1)\alpha_j = \rho'(1) \neq 0$$

since the LMM is convergent and therefore has a simple root at $z = 1$. Since ρ_i' is continuous, there is a neighborhood N_0 of zero in which $\rho_i' \neq 0$. Thus for $\theta \in N_0$, $\theta \neq 0$, we have $\rho_i(\theta) \neq 0$. QED.

Lemma 2.2-3. For any consistent LMM

$$\lim_{\theta \to 0} \frac{\rho_r(\theta)}{\rho_i(\theta)} = 0.$$

Proof:

$$\lim_{\theta\to 0} \frac{\rho_r(\theta)}{\rho_i(\theta)} = \lim_{\theta\to 0} \frac{\sum_{j=-1}^{k-1} \alpha_j \cos (k-j-1)\theta}{\sum_{j=-1}^{k-1} \alpha_j \sin (k-j-1)\theta}$$

$$= \lim_{\theta\to 0} \frac{\sum_{j=-1}^{k-1} (j+1-k)\alpha_j \sin (k-j-1)\theta}{\sum_{j=-1}^{k-1} (k-j-1)\alpha_j \cos (k-j-1)\theta} = 0$$

since consistency insures

$$\sum_{j=-1}^{k-1} (k-j-1)\alpha_j \neq 0.$$

QED.

Theorem 2.2-4. For any convergent LMM

$$\lim_{\theta\to 0} \alpha_\theta = \Pi/2.$$

Proof:

$$\lim_{\theta\to 0} \alpha_\theta = \lim_{\theta\to 0} \cot^{-1} \left(\frac{\rho_r\sigma_r + \rho_i\sigma_i}{\rho_r\sigma_i - \rho_i\sigma_r} \right)$$

$$= \cot^{-1} \lim_{\theta\to 0} \frac{(\rho_r/\rho_i)\sigma_r + \sigma_i}{(\rho_r/\rho_i)\sigma_i - \sigma_r} = \cot^{-1} 0 = \Pi/2.$$

since $\sigma(1) \neq 0$. QED.

Thus we see the boundary locus of the region of absolute stability always leaves the origin along the imaginary axis. Indeed then, if the non-principal roots of ρ are strictly inside the unit circle, there will be a non-vanishing interval of absolute stability to the left of the origin and there will be a non-null region of absolute stability in the left half plane.

For $\theta = 0, \Pi$, since ρ and σ are real polynomials, the boundary locus points will be real. At these values of θ we can compute the boundary locus points in general.

Theorem 2.2-5. The boundary locus points at $\theta = 0, \Pi$ for the region of absolute stability of any convergent LMM are the origin and $\rho(-1)/\sigma(-1)$, respectively.

Proof: For $\theta = 0$, convergence provides a simple root of $\rho(z)$ at $z = 1$. Consistency implies $\rho'(1) = \sigma(1)$, thus $\sigma(1) = 0$ forces a multiple root of ρ at $z = 1$. We conclude the origin is the boundary locus point at $\theta = 0$.

For $\theta = \Pi$ we have

$$\lim_{\theta \to \Pi} \operatorname{Re}\left[- \frac{\rho(e^{i\theta})}{\sigma(e^{i\theta})} \right] = \lim_{\theta \to \Pi} \frac{\rho_r \sigma_r + \rho_i \sigma_i}{\sigma_r^2 + \sigma_i^2} = \frac{\rho_r(\Pi)\sigma_r(\Pi)}{[\sigma_r(\Pi)]^2} = \frac{\rho(-1)}{\sigma(-1)}.$$

QED.

There may be other values of θ for which the boundary locus points are real. For those LMM of interest to the investigation reported in Chapters IV and V we characterize real boundary points in the theorem following. First we define a desired property.

Definition: A LMM is strictly zero-stable if the associated polynomial $\rho(z)$ has a simple root at $z = 1$ and all remaining roots interior to the unit circle.

This definition is clearly stronger than the zero-stability defined in Chapter I.

Theorem 2.2-6. Let L be a strictly zero-stable LMM for which no roots of $\sigma(z)$ have unit magnitude. Then the boundary locus points of the region of absolute stability are real if and only if $\rho(e^{i\theta})$ and $\sigma(e^{i\theta})$ lie on the same line through the origin.

Proof: When $\theta = 0$ the result is clear since both $\rho(1)$ and $\sigma(1)$ are real.

In the remainder of the proof we assume $\theta \neq 0$. The boundary locus points are real $\overset{\rightarrow}{\leftarrow}$

$$\text{Im}\left[-\frac{\rho(e^{i\theta})}{\sigma(e^{i\theta})}\right] = 0 \overset{\rightarrow}{\leftarrow} \frac{\rho_i\sigma_r - \rho_r\sigma_i}{\sigma_r^2 + \sigma_i^2} = 0 \overset{\rightarrow}{\leftarrow} \rho_i\sigma_r = \rho_r\sigma_i.$$

Suppose $\rho_i(\theta) = 0$. If the boundary locus points are real then $\rho_r(\theta)\sigma_i(\theta) = 0$. Since $\theta \neq 0$ and L is strictly zero-stable $\rho_r(\theta) \neq 0$ so that $\sigma_i(\theta) = 0$. Then both $\rho(e^{i\theta})$ and $\sigma(e^{i\theta})$ are real and lie on the same line through the origin. Conversely, if both $\rho(e^{i\theta})$ and $\sigma(e^{i\theta})$ are on the same line through the origin and $\rho_i(\theta) = 0$, both $\rho(e^{i\theta})$ and $\sigma(e^{i\theta})$ are real and the boundary locus point is real.

Suppose $\sigma_i(\theta) = 0$. If the boundary locus points are real then $\rho_i(\theta)\sigma_r(\theta) = 0$. Since $\sigma_r(\theta) = 0$ would yield a root of σ with unit magnitude we conclude $\rho_i(\theta) = 0$. Thus both $\rho(e^{i\theta})$ and

$\sigma(e^{i\theta})$ are real and lie on the same line through the origin. Conversely, if both $\rho(e^{i\theta})$ and $\sigma(e^{i\theta})$ lie on the same line through the origin and $\sigma_i(\theta) = 0$ then $\rho_i(\theta) = 0$ and both $\rho(e^{i\theta})$ and $\sigma(e^{i\theta})$ are real resulting in a real boundary locus point.

If both $\rho_i(\theta)$ and $\sigma_i(\theta)$ are non-zero by the remarks at the beginning of the proof we have the boundary locus point at θ real if and only if

$$\frac{\rho_r(\theta)}{\rho_i(\theta)} = \frac{\sigma_r(\theta)}{\sigma_i(\theta)} .$$

These are equivalent to $\rho(e^{i\theta})$ and $\sigma(e^{i\theta})$ lying on the same line through the origin. QED.

2.3 Stability for Stiff LMM

When $n > 1$ in problem (1.1-1) the eigenvalues of the system can be of vastly different magnitudes within the interval of integration. When this situation occurs stability constraints on the maximum stepsize permitted may be dictated throughout the interval by eigenvalues whose contribution to the solution of the system becomes negligible after a time. Often the stability constraint limits the stepsize so severely that the roundoff errors and computation time involved are overwhelming. It would be nice when this occurs to have a region of absolute stability extending to infinity to allow us to determine stepsize on the basis of accuracy requirements alone. This is the essence of the stiff problem and its solution.

We wish then to find LMM with infinite regions of absolute stability in the left half plane without sacrificing accuracy around the origin. First we review the more important definitions from the literature characterizing properties of such methods. One such definition has been provided by Gear [9] and describes stiffly stable LMM.

Definition: A LMM is stiffly stable if in the region R_1 ($\mathrm{Re}\, q \leq D$) it is absolutely stable and in R_2 ($D < \mathrm{Re}\, q < \alpha$, $|\mathrm{Im} q| < \theta$) it is accurate.

The region of accuracy referred to here is not precisely defined and therefore this definition is not entirely workable. Other definitions focus only on the infinite region of absolute stability and omit any measure for a region of accuracy about the origin. The following definition provides a generally unobtainable standard to which methods may be compared. It was introduced by Dahlquist in 1963.

Definition: A LMM is called A-stable if the region of absolute stability includes the open left half plane.

We follow here with a somewhat discouraging result also due to Dahlquist [6].

Theorem 2.3-1. The order of an A-stable LMM cannot exceed two.

This result naturally led to less severe definitions. Widlund [41] introduced the concept of A(α)-stability.

Definition: A LMM is A(α)-stable, $\alpha\varepsilon(0, \Pi/2)$ if the region of absolute stability includes the wedge

$$S_\alpha = \{z| \ |Arg(-z)| < \alpha, \ z \neq 0\}.$$

(Here Arg $z\varepsilon(-\Pi, \Pi]\forall \ z\varepsilon C$).

A method is A($\Pi/2$)-stable if it is A(α)-stable for all $\alpha\varepsilon(0, \Pi/2)$ and A(0)-stable if it is A(α)-stable for some (sufficiently small) $\alpha\varepsilon(0, \Pi/2)$.

Cryer [4] completed the spectrum of A(α)-stability with his definition of A_0-stability.

Definition: A LMM is A_0-stable if its region of absolute stability includes the negative real axis.

Clearly then for systems with real eigenvalues we need A_0-stability and for complex eigenvalues we need A(α)-stability where the α is determined by the eigenvalues of the system we are solving. Cryer [4] gives an example of a LMM which is A_0-stable but not A(0)-stable. There exists then a series of strict inclusions for the stability classes of LMM from A-stability through A_0-stability.

Cryer [4] shows that the Adams-Moulton formulae of order $k \geq 2$ are not A_0-stable but he does show that there exists A_0-stable methods of arbitrarily high order. The backward differentiation formulae implemented by Gear are not stiffly stable for order $k \geq 7$. Our investigation of LMM for use on stiff systems is restricted to implicit methods as we shall see presently. First a variation of a result from Rodabaugh [34].

Lemma 2.3.2. Let $\lambda_i(a)$, $i = -1, \cdots, m$ be continuous functions of a on the interval $(-\infty, b)$ with

$$\lim_{a\to-\infty} \lambda_{-1}(a) = 0.$$

Assume that for some i,

$$\lim_{a\to-\infty} \lambda_i(a) \neq 0.$$

Then for any $x > 0$ there exists an N such that, if $a < N$ then

$$\sum_{i=-1}^{m} \lambda_i(a) r^{m-i} = 0 \qquad (2.3\text{-}1)$$

has a root with absolute value larger than x.

Proof: Let $j = \min \{i \mid i \geq 0 \text{ and } \lim_{a\to-\infty} \lambda_i(a) \neq 0\}$. The roots of (2.3-1) are identical with those of

$$r^{m+1} + \sum_{i=0}^{m} (\lambda_i(a)/\lambda_{-1}(a)) r^{m-i} = 0 \qquad (2.3\text{-}2)$$

provided $\lambda_{-1}(a) \neq 0$. Since (2.3-2) is monic, its coefficients $(\lambda_i(a)/\lambda_{-1}(a))$ can be expressed by the symmetric polynomials as polynomial functions of its roots. However, as a tends to negative infinity, $\lambda_j(a)/\lambda_{-1}(a)$ increases without bound so at least one of the roots is unbounded. QED.

Theorem 2.3-3. Any explicit LMM is not A_0-stable.

Proof: Let $\Pi(z, q) = \rho(z) + q\sigma(z)$ be the characteristic polynomial of an explicit LMM where ρ and σ are of degree k and k-1, respectively. The polynomial $1/q\ \Pi(z, q)$ has the same roots as $\Pi(z, q)$. We need only consider real q and the coefficient of z^k

in 1/q $\Pi(z, q)$ as q approaches negative infinity. The previous lemma guarantees an N such that $\Pi(z, q)$ has roots exceeding unit magnitude for $q < N$. QED.

Jeltsch [20] finds that Cryer's [4] methods are not only A_0-stable but also A(0)-stable. Hence there exists A(0)-stable methods of arbitrarily high order. Gupta [14] has found A(α)-stable methods through order 12 with α exceeding 70°, however, the truncation error gets extremely large on the higher order methods.

Another useful definition is provided by Odeh and Linger [30].

Definition: A LMM is A_∞-stable if it is absolutely stable in a neighborhood of infinity on the complex q-sphere. We take the following theorem from Marden [29] without proof.

Theorem 2.3-4 (Rouché). If P(z) and Q(z) are analytic interior to a simple closed Jordan curve C and if they are continuous on C and $|P(z)| < |Q(z)|$ for $z \in C$, then the function $F(z) = P(z) + Q(z)$ has the same number of zeros interior to C as does Q(z).

The following lemma is based on Rouché's Theorem and is useful in providing insight into the concept of A_∞-stability. In the remainder of this section let L be an implicit zero-stable LMM with characteristic polynomial $\Pi(z, q) = \rho(z) + q\sigma(z)$ and $C = \{z \mid |z| = 1\}$.

Lemma 2.3-5. Consider L as described above and suppose none of the roots of σ lie on C. Then there exists a real number S such that for $|q| > S$, $\Pi(z, q)$ and σ(z) have the same number of zeros interior to C.

Proof: Let

$$S = \sup_{z \in C} \left| \frac{\rho(z)}{\sigma(z)} \right|.$$

Then S is finite and if $|q| > S$ we have $|q\sigma(z)| > |\rho(z)|$ for $z \in C$. By Rouché's Theorem we know that $\Pi(z, q)$ and $q\sigma(z)$ have the same number of zeros interior to C. Since the roots of $q\sigma(z)$ and $\sigma(z)$ are identical and since the degrees of Π and σ are the same, we have $\Pi(z, q)$ and $\sigma(z)$ with the same number of zeros interior to C. QED.

Now we prove a simple and useful characterization of A_∞-stable methods.

Theorem 2.3-6. L is A_∞-stable if and only if all the roots of $\sigma(z)$ lie inside C.

Proof: Let the degrees of Π be k and let S_i and $z_i(q)$, $i = 1, \cdots, k$ be the roots of $\sigma(z)$ and $\Pi(z, q)$, respectively, where

$$\lim_{q \to \infty} z_i(q) = S_i.$$

If L is A_∞-stable there is a neighborhood of $q = \infty$ in which $|z_i(q)| < 1$ so that clearly $|S_i| < 1$, $i = 1, \cdots, k$.

If $|S_i| < 1$, $i = 1, \cdots, k$ then the previous lemma provides A_∞-stability immediately. QED.

The result which follows provides a relationship between A_0-stabilizy and A_∞-stability.

Theorem 2.3-7. If L is A_0-stable and none of the roots of $\sigma(z)$ lie on C, then L is A_∞-stable.

Proof: Suppose one of the roots of σ, say S_j has magnitude greater than one. The previous lemma provides a real number S with a root of $\Pi(z, q_0)$ outside C for all q_0 with $|q_0| > S$. We therefore contradict A_0-stability and we conclude all roots of σ lie inside C and hence L is A_∞-stable. QED.

The last definition we look at from the literature is from Gupta [13] and combines features of Gear's stiff stability and the $A(\alpha)$-stability of Widlund.

Definition: A LMM is said to be $A(\alpha, D)$-stable, $\alpha \in (0, \Pi/2)$ if the region of absolute stability includes all q with $|\text{Arg}(-q)| < \alpha$, $q \neq 0$ and all q with $\text{Re } q \leq D$.

Of all the definitions reviewed above, the only one to define a region of accuracy about the origin was Gear's stiff stability definition. As mentioned before the reference to accuracy in that definition is vague. This has led to different interpretations of what is meant by the region of accuracy about the origin for example see Jeltsch [20, p. 9] and Gupta [15, p. 492]. The motivation for developing stiffly stable methods was to allow us to determine stepsize on the basis of accuracy constraints alone. To that end we do need a description of the infinite region of stability as provided by the foregoing definitions. However, we cannot overlook then the parallel need to provide a description of the region of accuracy with equivalent precision. Absolute stability as mentioned before provides a measure of accuracy which is too lax in the left half plane and too rigid in the right half plane. We must be concerned with digits of precision when performing computations on a

digital computer. Actually digits of precision are a measure of relative error. For these reasons then it seems natural to continue our concern about relative error with a measure of relative stability which we include in the definition given below.

Definition: A LMM is said to be $A(\alpha, r)$-stable if the method is $A(\alpha)$-stable with regard to its region of absolute stability and is relatively stable within the disk of radius r about the origin.

It is this definition which we implement in our search for stiffly stable LMM. A method which is $A(\alpha, r)$-stable will be relatively robust depending upon the size of α and r. By that we mean it should give good results on a large variety of problems--both stiff and non-stiff.

CHAPTER III

ERROR ANALYSIS

3.1 Introduction

Error in our approximation to the solution of (1.1-1) occurs because we base our recursive calculations on inaccurate values, because our recursive model does not accurately represent the problem, and because our calculations themselves are not always precise. Also, the recursive nature of our model inherently propagates existing error. Controlling this propagation of error is the aim of stability constraints discussed in Chapter II. Any complete and careful analysis of cumulative error must consider at least these sources of error.

3.2 A Global Bound

Henrici [16, section 5.3-4] investigates this cumulative error and arrives at the bound given below for the solution of (1.1-1) by a member of C(p, k). If

$$|\alpha_{-1}^{-1}\,\beta_{-1}|hL < 1$$

we have the global error e_n satisfying

$$|e_n| \leq \Gamma^*[A\delta k + (x_n-a)(k_1h^q + Gyh^p)]\exp[(x_n-a)\,L\Gamma^*B] \quad (3.1\text{-}1)$$

where the following definitions are used.

$$A = \sum_{i=-1}^{k-1} |\alpha_i|, \quad B = \sum_{i=-1}^{k-1} |\beta_i|,$$

L is the Lipschitz constant for the function f, δ is the maximal starting value error,

$$\Gamma = \sup \{|\lambda_i|\}_{i=0}^{\infty}$$

where

$$\frac{1}{\hat{\rho}(z)} = \sum_{i=0}^{\infty} \lambda_i z^i$$

and

$$\hat{\rho}(z) = \sum_{i=-1}^{k-1} \alpha_i z^{i+1}$$

(Γ is shown to be finite in Henrici [16, p. 242] for zero-stable methods.)

$$\Gamma^* = \frac{\Gamma}{1-h|\alpha_{-1}^{-1}\beta_{-1}|L},$$

h is the stepsize, $x_n \varepsilon [a, b]$, $k_1 h^{q+1}$ is a bound for the magnitude of the roundoff error committed at any step of the integration process,

$$G = \int_0^k |G(s)| ds$$

where

$$G(s) = \sum_{j=-1}^{k-1} \alpha_{k-j-2}(j-s)_+^p - p\beta_{k-j-2}(j-s)_+^{p-1}$$

and

$$(j-s)_+ = \max\{0, j-s\},$$

and

$$y = \max_{x\in[a,b]} |y^{(p+1)}(x)|.$$

Error bounds generally have the burden of coping with worst case examples and therefore many times are forced to give outrageously large bounds on simple problems. This severely damages their practical value.

Consider for example the bound (3.1-1) applied to the solution of

$$y' = \lambda y,\ y(0) = 1,\ x\in[0, 1] \qquad (3.1\text{-}2)$$

using the second order backwards difference formula

$$y_{n+1} = 4/3\ y_n - 1/3\ y_{n-1} + 2/3\ hf(x_{n+1}, y_{n+1}) \qquad (3.1\text{-}3)$$

with stepsize h. Then

$$\rho(z) = -z^2 + 4/3\ z - 1/3$$

so that

$$\frac{1}{\hat{\rho}(z)} = \frac{-3}{(z-1)(z-3)} .$$

For $|z| < 1$ we find the Laurent expansion

$$\frac{1}{\hat{\rho}(z)} = \sum_{i=0}^{\infty} \left[\frac{1-3^{i+1}}{2\cdot 3^{i}}\right] z^{i} .$$

Hence

$$\Gamma = \sup \{|\lambda_i|\}_{i=0}^{\infty} = \sup \left\{\left|\frac{1-3^{i+1}}{2\cdot 3^{i}}\right|\right\}_{i=0}^{\infty} = 3/2.$$

In general for the method (3.1-3) we see that (3.1-1) reduces to

$$|e_n| \leq \frac{1}{6-4hL} \{48\delta + (x_n-a)(9k_1h^q + 2yh^2)\} \exp\left[\frac{3L(x_n-a)}{3-2hL}\right] \quad (3.1\text{-}4)$$

and is valid as long as $hL < 3/2$.

In this example we then have the bound as

$$|e_n| \leq \frac{48\delta + 9n\, k_1h^{q+1} + 2nL^3h^3}{6-4hL} \exp\left[\frac{3nhL}{3-2hL}\right]. \quad (3.1\text{-}5)$$

If $\lambda = -100$ and $h = 0.01$ then $hL = 1 < 3/2$ so that this expression is valid. We compute this bound on the error after 25 steps and in doing so it is of negligible consequence to omit the positive contributions of starting error and roundoff error. In this case

$|e_{25}| \leq 25 \exp(75) \approx 9.33 \cdot 10^{33}$ when the true solution $y(x_{25}) \approx .138 \cdot 10^{-10}$. Clearly the information provided by this bound from a practical point of view is totally useless. When error bounds behave in this way we can sometimes get better results from the use of error estimates [9, p. 14]. However, as with the example given by Gear [9, p. 16], error estimates can give smaller numbers than the actual error.

From a practical point of view it would be more useful if the expressions for roundoff error and starting error were respectively dependent upon machine precision and accuracy of the starting procedure. The bound (3.1-1) is theoretically satisfying from the viewpoint that it does approach zero with h. It would be pleasing to see the effect of stability properties represented in the bound. In the next section we consider a representation of the global error which is more useful for certain practical and theoretical considerations.

3.3 An Approximate Error Representation

As in section 1.2, if we assume our solution $y(x)$ has sufficient derivatives, the local truncation error can be expressed as

$$L(y(x_{n+1}), h) = \sum_{i=-1}^{k-1} (\alpha_i y_{n-1} + h\beta_i y'_{n-i})$$

$$= C_{p+1} h^{p+1} y^{(p+1)}(x_{n+1}) + O(h^{p+2}) \qquad (3.3\text{-}1)$$

for a member of C(p, k) where this notation is taken from Henrici [16, p. 220]. L, as defined above, may be regarded as a linear operator on any differentiable function $y(x)$, and if y possesses sufficient derivatives its value can be given as the right side of (3.3-1). Following the investigation of Henrici for arbitrary operators of the form (3.3-1), others [27, 33] have pointed out that in all cases

$$|L(y(x_{n+1}), h)| \leq \frac{1}{p!} h^{p+1} Gy, \tag{3.3-2}$$

where G and y are defined as in the previous section. G(s), also defined in the previous section, is called the influence function. In those cases where G(s) is of the same sign throughout the interval [-1, k-1], we have

$$L(y(x_{n+1}), h) = C_{p+1} h^{p+1} y^{(p+1)}(\xi). \tag{3.3-3}$$

for $\xi \varepsilon (x_{n-k+1}, x_{n+1})$. G(s) does not change sign on [-1, k-1] for many commonly used methods including Adam's formulae and the backwards difference formulae of Gear with order below six.

We now develop an approximate representation of the global error which occurs in the solution of $y' = \lambda y$ by a member of C(p, k). Let $y(x_n)$ and y_n be the true and approximate solution, respectively. Let the local truncation error and roundoff error occurring in the calculation of y_n be T_n and R_n respectively. Then $T_n = L(y(x_n), h)$ and

$$R_n = \sum_{i=-1}^{k-1} (\alpha_i + h\lambda\beta_i) y_{n-1-i}.$$

If we define $e_n = y(x_n) - y_n$ we have

$$T_{n+1} - R_{n+1} = \sum_{i=-1}^{k-1} (\alpha_i + h\lambda\beta_i) y(x_{n-i}) - \sum_{i=-1}^{k-1} (\alpha_i + h\lambda\beta_i) y_{n-i}$$

$$= \sum_{i=-1}^{k-1} (\alpha_i + h\lambda\beta_i) e_{n-i}. \qquad (3.3\text{-}4)$$

If, as in Lambert [27, p. 68], we assume $T_{n+1} - R_{n+1} = \varphi$, a constant, we have

$$e_n = \sum_{i=0}^{k} d_i z_i^n + \frac{\varphi}{h\lambda\beta_i} \qquad (3.3\text{-}5)$$

where z_i are the roots of the characteristic polynomial $\Pi(z, h\lambda)$. This results as the solution to the non-homogeneous constant coefficient difference equation [27, p. 8]. The form in which we have written (3.3-5) assumes the roots z_i are distinct.

Admittedly, the form of the second term in (3.3-5) will not be precisely as represented if $T_{n+1} - R_{n+1}$ is not constant. However, the first term keeps the form given regardless of the form φ takes. The first term shows the interplay between the stability properties of a method and perturbations such as roundoff and truncation error introduced at each step.

When solving a problem such as (1.1-1) with values of $h\lambda$ inside of absolute stability but outside the region of relative stability, the truncation error will approach zero but not as rapidly as the true solution. The calculation then is forced to continue until the error comes within accuracy constraints even though the actual solution may have long been satisfactorily small. The expense of this continued calculation offsets and conceivably could exceed the expense of using a smaller stepsize to bring $h\lambda$ within the region of relative stability. This is why relative stability can be important even in the left half plane. Examples of this were obtained in computer runs using 35-digit precision. Solution of (3.1-2) with $\lambda = 100$ and $\lambda = -1$ using the method (3.1-3) and $h = 0.01$ gave relative error of $2.53 \cdot 10^{8}$ and $3.3 \cdot 10^{-5}$, respectively, after 100 steps.

Inside the region of relative stability the second term in (3.3-5) usually dominates the global error. The problem independent part of the local and global truncation error constants are

$$\frac{C_{p+1}}{\alpha_{-1}} \text{ and } \frac{C_{p+1}}{\sigma(1)},$$

respectively. If we use sufficient machine precision to insure the roundoff error is negligible compared to the truncation error, then in (3.3-5) ϕ is dominated by the local truncation error

$$\frac{C_{p+1}}{\alpha_{-1}}.$$

This makes it apparent why the problem independent part of the local truncation error differs from the global truncation error by a factor of

$$\frac{\alpha_{-1}}{\sigma(1)} .$$

This observation is supported by computer runs, again in 35-digit precision, where in the solution of (3.1-2) with $\lambda = -1$ using a variety of LMM and stepsizes the ratio of local to global truncation error was consistently

$$\frac{\alpha_{-1}}{\sigma(1)}$$

within approximately three significant digits. It is important to note here that there is an entire subclass of $C(p, k)$ which have local truncation error constants larger than their global truncation error constants, and conversely. This results since the determination of C_{p+1} and $\sigma(1)$ are independent in the solution of the system of equations (1.2-3) which defines the members of $C(p, k)$. Evaluation of the relative merits of different LMM need to take this into account. In the investigation reported in Chapter V the term

$$\frac{C_{p+1}}{\sigma(1)}$$

is used as the measure of accuracy for comparison of LMM.

CHAPTER IV

PREDICTOR SEARCH

4.1 Statement of and Approach to the Problem

Predictor methods commonly suffer smaller stability regions and larger truncation error constants than implicit methods [27, p. 84]. On the other hand, implicit methods suffer the difficulty of finding the solution to the implicit relation defining the forward solution approximation at each step of the integration. The solution to these problems is normally accomplished by pairing a predictor with a corrector to form a predictor-corrector (P-C) algorithm. In the solution of non-stiff problems, if $\{y_i \mid i = 0, \cdots, N\}$ is our discrete model of the solution to (1.1-1), the predictor is used to compute a first approximation to forward values y_{n+1}. This predicted value is then used as $y_{n+1}^{[1]}$ in (1.2-5) to start the recursive evaluation of the sequence

$$\{y_{n+1}^{[s]}\}_{s=1}^{\infty} .$$

This sequence may be evaluated as far out as desired although convergence within desired accuracy constraints is usually very fast. For example the popular code DIFSUB by Gear [10] restricts $s \leq 4$. In the solution of stiff problems, the condition required for convergence of the corrector iteration,

$$|h\beta_{-1}\alpha_{-1}^{-1}|L < 1$$

where L is a Lipschitz constant for $f(x, y)$ with respect to y, forces us to discard corrector iteration in favor of a Newton iteration [27, p. 239]. When Newton's method is applied to (1.2-5) for the solution y_{n+1}, the restriction on h required for convergence is usually much more relaxed than that imposed by

$$|h\beta_{-1}\alpha_{-1}^{-1}|L < 1.$$

But even in stiff problems, predictors are used to provide accurate initial estimates for the Newton iteration.

If predictors with significantly improved stability characteristics exist, at least two interesting possibilities arise. Those possibilities are improved P-C algorithms and codes using predictors exclusively. In a P-C algorithm as described in the previous paragraph, unless the corrector iteration is carried out to convergence, the stability characteristics of the P-C algorithm are not those of the corrector alone [1]. The fewer the number of corrector iterations performed, the less heavily the stability of the P-C algorithm depends on the corrector. Since, as noted above, in practice few iterations are actually carried out it leads us to believe that it may be important to use a predictor with as large a stability region as possible.

The other possibility of using predictors alone for non-stiff problems is very attractive in all cases and occasionally the use of

explicit methods is necessary. Among the advantages to be gained are economy, simplicity, and versatility. Economy is realized with regard to speed of computation, programming efforts in developing a reliable code, and machine storage. For most non-trivial problems, by far the most expensive computation performed in finding the model $\{y_i\}_0^N$, is that of function evaluations for the derivative approximations y'_{n+1}. Predictor algorithms reduce this to one function evaluation per step whereas P-C algorithms require from two to four. Elimination of the corrector iteration and the associated coding greatly reduces the programming effort. It also simplifies considerations such as stability analysis. For example see Lambert [27, p. 97] for a definition of the characteristic polynomial of a P-C method. A proper stability analysis must examine the roots of the characteristic polynomial which we see is much more complicated than for a predictor alone. The reduced coding and storage required of predictor algorithms render such codes' implementation within the capabilities of smaller machines, even hand-held calculators when the problem is not large.

Gear [11] gives real-time integration as an example of where implicit methods cannot be used in the usual sense. Here again the use of predictors would be natural were it not for the smaller stability regions of these explicit methods.

It is true that normally predictor methods also have larger truncation error constants than correctors but this can always be more than compensated for by using reduced steplength. For example in comparing the truncation error constants of the Adams-Bashforth to Adams-Moulton methods, the stepsize reduction factor required to

equilibrate the truncation error constants for a given order increases from 0.447 for second order to 0.671 for ninth order. It is the smaller stability regions which create the greatest drawback for the exclusive use of predictors in solving certain problems (1.1-1).

To date no comprehensive search has been conducted for more stable predictors. Those searches existing in the literature have had limited specific goals, for example see [24, 3]. The purpose of the investigation reported in this chapter then is to determine how far the stability regions of predictor methods can be extended. These methods may find utility by incorporation into P-C algorithms or in a code which employs explicit methods exclusively. (A related question, not investigated here, is whether or not it is the case that P-C algorithms with better stability properties are always the result of combining separate predictors and correctors, each with good stability properties.)

A guide as to the class of predictors in which we may expect favorable results is provided by Henrici [16, section 5.2-8]. The theorem is given here without proof.

Theorem 4.1-1. The order of a zero-stable LMM whose stepnumber is k cannot exceed k+1 when k is odd or k+2 when k is even.

As discussed in Lambert [27, p. 67], those zero-stable methods of the class P(k+2, k) when k is even have no interval of absolute stability. We are thus restricted to investigation of classes P(n, m) where $n \leq m + 1$. We choose the class P(n, n) because of the additional degree of freedom permitted over P(n, n-1) and

because many codes already in existence have implemented the Adams-Bashforth methods which lie in this class.

We use the relative stability region of a LMM as our measure of stability. We could use absolute stability regions for such a measure, however this would provide us no knowledge of the behavior of our approximate solution to (1.1-1) for q in a neighborhood of the origin. We define the radius of relative stability to be the radius of the largest circle centered at the origin and contained in the region of relative stability.

There are several definitions of relative stability existing in the literature [27, p. 68], and it is important to distinguish which definition is used when making comparisons of stability regions from different sources. Even though for small q we have the relationship (2.1-1) there can be large differences for q near zero. In our investigation we use the definition given in section 2.2 which is taken from Lambert [27]. Another approach to defining relative stability is to compare the magnitude of the roots of the characteristic polynomial to the magnitude of exp (q) rather than to the magnitude of the principal root. For example, Crane and Klopfenstein require $|r_s| \leq \exp(q)$, $s = 1, 2, \cdots, k$ and that roots of magnitude exp (q) be simple. For comparison, consider the second order backwards difference formula (3.1-3). This method has characteristic polynomial $\Pi(z, q) = 2/3(q-1)z^2 + 4/3\ z - 1/3$. For the sake of simplicity, we restrict our comparison to the interval of relative stability which is the real values of q for which the method is relatively stable. Lambert's and Crane and Klopfenstein's definition yield the values

$(-1/2, 3/2) \cup (3/2, \infty)$ and $(a, 0]$, respectively, where $a \approx -0.75262$. Other methods tested indicate possibly for Crane and Klopfenstein's definition the radius of the complex region of relative stability is always zero. For example this is the case when the boundary of the region of absolute stability intersects the imaginary axis only at the origin. This follows since for q imaginary, $|\exp(q)| = 1$, and the condition of relative and absolute stability coincide. Also if y_n and y_n are the true and computed solutions at a point x_n, then the latter definition restrains $|y_n| = |y_n| - \varepsilon$, for $\varepsilon > 0$. The magnitude of ε and not its sign determine the acceptability of y_n. For these reasons we find the latter definition unsuitable. Lambert's definition ties the growth of the relative error to the size of the principal root, however, for q not large and especially for higher order methods (2.1-1) indicates this is not a lax requirement. Some definitions of relative stability which compare magnitudes of the roots of the characteristic polynomial to that of the principal root do not require strict inequality, see for example Ralston [33, p. 177]. However, strict inequality more easily lends itself to computer implementation.

We need then to find a means of evaluating the radius of relative stability defined above. An expression for the boundary of the region of relative stability cannot be found as with absolute stability. It is necessary to actually find the roots of the characteristic polynomial. We define a function introduced by Rodabaugh [34] and called the critical difference function. Given a LMM with characteristic polynomial $\Pi(z, q)$, we define

$$C(q) = |z_1(q)| - \max \{|z_S(q)|\}_{S=2}^{k}$$

where $z_S(q)$, $S = 1, 2, \cdots, k$ are the roots of $\Pi(z, q)$ and $z_1(q)$ is the principal root. In our search for stable predictors we require $C(0) > 0$, or strict zero-stability. We are thus guaranteed a neighborhood of stability about the origin by the continuity of the critical difference function.

To determine the radius of relative stability we need then to locate the zero of $C(q)$ nearest the origin. The radius will be equal to the modulus of this zero.

We do this numerically by taking a finite set of rays through the origin, finding the zero of $C(q)$ nearest the origin on each ray in this set, and approximating the radius as the smallest modulus possessed by the resulting zeros. However, to locate the zero of $C(q)$ nearest the origin on a given ray we must be able to evaluate $C(q)$ and this entails tracking the principal root.

To track the principal root along a ray, q is given a fixed argument and its modulus is incremented successively by a small amount. We know $z_1(0) = 1$ and if q_n and q_{n+1} are successively incremented values of q on a given ray, then $z_1(q_{n+1})$ is assigned to be the root of $\Pi(z_1\ q_{n+1})$ nearest $z_1(q_n)$. Thus starting at the origin $C(q)$ is evaluated at the successive increments of q as we track the principal root. The first increment at which $C(q)$ is negative identifies an interval which contains the first zero of $C(q)$. A bisection routine is then used to more accurately identify the zero. It is conceivable that the first zero of $C(q)$ could be

missed in this way, however, it would be necessary for it to be missed on several rays in order for the output value of the radius to be severely affected. This is highly unlikely but we should remain aware of the possibility.

Evaluation of the radius of relative stability can be costly and rather complex. It is possible that much research into relative stability characteristics possessed by LMM has been inhibited by these factors. In addition, these factors may have partially motivated the alternative definitions which compare the roots of $\Pi(z, q)$ to exp (q) rather than to the principal root. There are several considerations which greatly reduce the cost factor involved.

These considerations are based on properties of the characteristic polynomial $\Pi(z, q)$ and the resulting boundary of consistent zero-stable methods. First of all, it suffices to check for zeros of C(q) when Im (q) $\geq$ 0, since Lemma 2.2-1 provides that the boundary of the region of relative stability is symmetric about the real axis. We also economize by restricting Re (q) $\leq$ 0 since in practice it is there that we find the most severe restrictions on the radius. Normally we find it sufficient to determine zeros of the function C(q) on 15 rays in the second quadrant, however, occasionally for verification we increase this number and extend our investigation into the first quadrant. Since we search only for the radius, we terminate our search for a zero of C(q) on a given ray once we have incremented the modulus of q beyond the modulus of any previously determined zero. Thus due to the shape of a typical boundary it is less time consuming to assign the first ray an argument of Π and proceed

backwards to an argument of $\pi/2$ rather than conversely. Finally, based on (2.1-1) we expect $\exp(q) \overset{\sim}{=} z_1(q)$ for q small. A computer analysis of this approximation has shown this to be of practical value as well. For a large number of methods tested through ninth order, when $|q| \leq 0.25$ the different $|\exp(q) - z_1(q)|$ was always at most on the order of 10^{-2}. In practice then we use the alternate means of identifying the principal root as that root of $\pi(z, q)$ nearest $\exp(q)$, for $|q| \leq 0.25$. In this range of moduli much larger increments are permitted. This feature alone has resulted in a cost-savings of approximately 80%. The savings here is greater when the radius of the method is small, for example in higher orders. It is difficult to assess the reduction in the cost of this investigation afforded by the combination of factors mentioned above, however, it is clear that they have vastly increased its feasibility from the cost perspective.

The search for stable predictors within the class P(n, n) is posed as an optimization problem. In order to do this we isolate a specific real quantity which is to be minimized. That quantity is the negative of the radius of relative stability. Clearly there are limitations incurred by selecting a specific objective such as we have. For example in the nature of the resulting coefficients and error constants of the optimal method. However, our results prove valid as inputs to decision processes concerning further development of numerical methods for approximating the solution of problems (1.1-1), and as numerical methods in their own right for some classes of problems.

Two packaged codes were incorporated into the program written to locate stable predictors. The optimization code used (ZCXPWLD) is based on Fletchers [8] version of Powell's [32] penalty function. It uses Powell's unconstrained optimization algorithm [31] to determine a local minimum through a sequence of unconstrained minimization problems. For the most part it was not necessary to use constraints.

In those cases we gained convergence more rapidly and the algorithm reduced to that described in [31]. Another code (ZCPOLY) employed in our search finds the zeros of a complex polynomial. This code is in the International Mathematical and Statistical Libraries (IMSL). It uses the Jenkins-Traub three stage variable shift iteration described in [21] and [22]. The remainder of the program was developed by this writer.

The domain of our object function for the class P(n, n) is p^{n-1}. This results from consideration of the system of equations (1.2-3) which determines the class P(n, n) as discussed in section 1.3. The only modification, to include notation, of that discussion applicable here is that for predictors $\varepsilon_{-1} = 0$. The flow of the program consists of taking a point $(\alpha_1, \cdots, \alpha_{n-1})$ in R^{n-1} and determining the coefficients of the corresponding LMM according to the procedures outlined in section 1.3. We then evaluate C(0). If $C(0) \leq 0$ the object function is assigned a zero value. Otherwise the radius is evaluated as previously described and the object function is assigned the negative of the radius. Based on this assigned function value,

the optimization algorithm calculates a new domain point in R^{n-1}. We then repeat this cycle until functional values converge within a prescribed tolerance. This optimization code perfor[illegible]well, usually converging within 100 function evaluations.

Essential for convergence is a starting point (method) with a positive radius. Good starting points are readily available from the literature, for example the Adams-Bashforth methods. We may gain additional starting values from consideration of a theorem from Chapter II. If we factor $\rho(z) = (z-1)\rho_S(z)$, then those polynomials $\rho(z)$ for which $\rho_S(z)$ is a Schur polynomial are associated with LMM of positive radius. Theorem 2.1-4 provides a characterization of the smallest convex set in R^n containing the coefficients of all Schur polynomials of degree n. For example we compute this characterization explicitly for information regarding starting values for the class P(5, 5). We have $\alpha_{-1} = -1$ and

$$\alpha_0 = 1 - \sum_{i=1}^{4} \alpha_i$$

so that $\rho(z) = (z-1)\ \rho_S(z)$, where

$$\rho_S(z) = \sum_{i=0}^{4} \left[\sum_{i=0}^{4} \alpha_j \right] z^{4-i}.$$

We apply Theorem 2.1-4 to the polynomial $\rho_S(z)$.

The matrix M_4 is given by

$$M_4 = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 \\ -4 & 2 & 0 & -2 & 4 \\ 6 & 0 & -2 & 0 & 6 \\ -4 & -2 & 0 & 2 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \qquad (4.1\text{-}1)$$

and

$$(a_0, a_1, a_2, a_3, a_4) =$$

$$(\alpha_4, \alpha_3 + \alpha_4, \alpha_2 + \alpha_3 + \alpha_4, \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4, 1).$$

The necessary condition in the Theorem is satisfied by requiring all elements of the product $M_4(a_0, \cdots, a_n)^T$ to be positive. This leaves us with the following systems of inequalities.

$$\begin{aligned} 1 - \alpha_1 - \alpha_3 &> 0 \\ 2 - \alpha_1 - \alpha_2 - 2\alpha_4 &> 0 \\ 3 - \alpha_2 - \alpha_3 + 2\alpha_4 &> 0 \\ 2 + \alpha_1 + \alpha_2 - 2\alpha_4 &> 0 \\ 1 + \alpha_1 + 2\alpha_2 + 3\alpha_3 + 4\alpha_4 &> 0. \end{aligned} \qquad (4.1\text{-}2)$$

4.2 Results and Comparisons

The results from the investigation described in the previous section are given here for orders four through nine. Since the Adams-Bashforth (A-B) methods are used nearly exclusively in popular codes employing explicit LMM, and since the K-th order A-B method is a member of P(K, K), we choose these methods as a standard for

comparison. The k^{th} order A-B method is characterized simply by

$$\rho(z) = z^{k-1} - z^k,$$

so it is the member of P(k, k) where all the parameters designated as free in section 1.3 have been set to zero. Classically, this has been regarded as a desirable property, and certainly one not possessed by the members of P(k, k) in general. It was thought necessary, for a general k^{th} order method, to store 2k values between integration steps and only k+1 for the A-B methods. For example this notion prevailed in limiting the search of Dill and Gear [7] for stiffly stable methods to this type of "minimum storage methods." However, Skeel [38] found that for P-C methods where a k^{th} order corrector is paired with a specific k^{th} order predictor it is necessary to store at most k+1 values, regardless of zero coefficients. If we apply these findings to codes employing predictors alone we see that any k-step method needs at most k values stored between integration steps. This removes storage economy as a relevant comparison.

As indicated in Table 4.2-1, we have found predictors with stability radii roughly three times that of the A-B method of corresponding order. Another valuable perspective is gained by noting the radius of any order A-B methods is roughly matched by our near-optimal method of two orders higher. Normally an increase in order means more accuracy at the expense of computation time induced by the smaller stepsize required of reduced higher order stability regions. As illustrated later in this section, we can leave

the stepsize the same and change from an A-B method to a near-optimal method of two orders higher.

Table 4.2-1.

Order	Adams-Bashforth		Near-Optimal		
	Radius	$C_{p+1}=C_{p+1}/\sigma(1)$	Radius	C_{p+1}	$C_{p+1}/\sigma(1)$
4	0.2146	0.3486	0.4706	0.3856	0.7711
5	0.1266	0.3297	0.3470	0.3666	1.3609
6	0.0731	0.3156	0.2395	0.3496	7.4542
7	0.0412	0.3042	0.1595	0.3259	2.1067
8	0.0226	0.2947	0.0745	0.3056	1.4813
9	0.0121	0.2870	0.0405	0.2957	2.4933

We placed no control on the expression for local or global truncation error constants. However, the local truncation error constant remained approximately the same as that for the A-B methods. The constant related to global truncation error, $C_{p+1}/\sigma(1)$, did increase moderately. This is a result of the observation that as we approach optimally stable methods, a root of $\sigma(z)$ goes to 1. It also appears there is no appreciable gain in stability when the constant $C_{p+1}/\sigma(1)$ is increased beyond a certain point. As shown by example later in this section, the increase of this constant for these near-optimal methods is not so severe as to damage their utility. Another study addressing the question of how far these stability regions can be extended for fixed error constants would be of interest.

Properties of these near-optimal predictors are given in Table 4.2-1. Rational coefficients for these methods are given in Appendix B. Note for A-B Methods, $\sigma(1) = 1$ so that $C_{p+1}/\sigma(1) = C_{p+1}$.

To illustrate the effectiveness of the near-optimal stability properties given in Table 4.2-1 and the necessity to measure relative stability we solve two example problems in 35-digit precision with A-B methods and with the near-optimal methods. The first problem considered is

$$y' = -16y,\ y(0) = 1,\ x\varepsilon[0,\ 3].$$

We use the A-B 5th order method on the 7th order near-optimal method, both with a stepsize of $h = 0.01$. Thus we have $h\lambda = 0.16$ which is inside both methods' region of absolute stability. However it is well outside the A-B 5th order relative stability region and near the boundary of relative stability for the 7th order near-optimal method. This problem gives us an example of what can happen when we forget about relative stability in the left half plane. If we had solved this problem with the A-B 5th order method and desired to follow the solution to this problem (possibly a component of a larger system) until the solution fell below 10^{-10} for example, we would find by continuing the calculation it is necessary to continue to step 825. Whereas in fact, and as detected by the relatively stable 7th order near-optimal method, we could terminate calculation after step 150. In this case and others like it, ignoring the relative stability characteristics of a method can be a very costly choice. The relative error for the A-B solution in Table 4.2-2 after 300 steps

is 1.6000Q+13 and for the near-optimal method is 5.3610Q-03. Note that the A-B solution is tending to zero, which evidences the methods' absolute stability at $h\lambda = -0.16$, but not nearly so fast as the solution itself. In fact it tends to zero so slowly that the calculated solution values are essentially meaningless.

Table 4.2-2.

Step No.	True Solution	A-B 5th Order		7th Order Near-Optimal	
		Calculated	Actual Error	Calculated	Actual Error
0	0.1000Q+01	0.1000Q+01	0.0	0.1000Q+01	0.0
50	0.3355Q-03	0.3357Q-03	-0.1988Q-06	0.3354Q-03	0.6280Q-07
100	0.1125Q-06	0.3001Q-06	-0.1875Q-06	0.1125Q-06	0.8515Q-10
150	0.3775Q-10	0.1108Q-06	-0.1108Q-06	0.3769Q-10	0.6116Q-13
200	0.1266Q-13	0.6541Q-07	-0.6541Q-07	0.1263Q-13	0.3401Q-16
250	0.4248Q-17	0.3862Q-07	-0.3862Q-07	0.4231Q-17	0.1686Q-19
300	0.1425Q-20	0.2280Q-07	-0.2280Q-07	0.1418Q-20	0.7640Q-23

The second example is taken to show simply that indeed there are values of $h\lambda$ outside the A-B methods' stability region but inside the near-optimal methods' stability region. Besides highlighting the difference in the stability regions this example provides credence in the method used to evaluate stability. The problem used is

$$y' = -y,\ y(0) = 1,\ x\varepsilon[0,\ 10].$$

It is solved by the 7th order A-B method and the 7th order near-optimal method given in Table 4.2-1. A stepsize of $h = 0.1$ was used

in both solutions resulting in $h\lambda = -0.1$ which, according to the table, is outside the A-B stability region but inside the near optimal stability region. The results are given below in Table 4.3-3.

Table 4.2-3.

Step No.	True Solution	A-B 7th Order		7th Order Near-Optimal	
		Calculated	Actual Error	Calculated	Actual Error
0	0.1000Q+01	0.1000Q+01	0.0	0.1000Q+01	0.0
25	0.8208Q-01	0.8208Q-01	0.5137Q-07	0.8208Q-01	0.4764Q-07
50	0.6738Q-02	0.6788Q-02	-0.4977Q-04	0.6738Q-02	0.1320Q-07
75	0.5531Q-03	-0.5459Q-01	0.5514Q-01	0.5531Q-03	0.1936Q-08
100	0.4540Q-04	0.6109Q+02	-0.6109Q-02	0.4540Q-04	0.2304Q-09

We see clearly that the A-B method is unstable for this problem.

Being able to increase the order by two without reducing the stepsize more than offsets the larger error constants possessed by the near-optimal methods. This can be done as long as we are inside the stability region and the stepsize does not exceed a certain size. For example, the part of the error dependent upon method and stepsize for the 7th order A-B method is $0.3042h^7$ and for the 9th order near-optimal method is $2.4933h^9$. Thus for $h < 0.3493$ the 9th order near-optimal method is more accurate than the 7th order A-B method. We compute these limiting values of h for all such comparisons in Table 4.2-4 below.

Table 4.2-4.

A-B Order	Near-Optimal Order	Limiting h-Value
4	6	0.2163
5	7	0.3956
6	8	0.4616
7	9	0.3493

We note all these limiting values of h are well outside normal stepsize ranges and that for h less than these values, increased accuracy will be realized by the higher order near-optimal methods. For example, in comparing the 6th order A-B method with the 8th order near-optimal method, the part of the error dependent upon stepsize and error constant for h = 0.1 is 0.3156Q-07 and 0.1481Q-08, respectively.

CHAPTER V

STIFFLY STABLE CORRECTOR SEARCH

5.1 Statement of and Approach to the Problem

In the past fifteen years much research has been devoted to the development of numerical methods for obtaining solutions to problems (1.1-1) of the class referred to as stiff problems. These problems and the difficulties they engender were described in section 2.1. Most popular codes which employ LMM for the solution of stiff problems depend exclusively on the backwards difference formulae (BDF) introduced by Gear. These methods suffer large regions of instability in the left half plane and are not A_0-stable for orders greater than six. This suggests much improvement is possible.

As mentioned in Chapter II, Grigorieff and Scholl [12] and Kong [25] have given constructive proofs which show the existence of $A(\alpha)$-stable methods of arbitrarily high order with α arbitrarily close to $\pi/2$. However, searches for methods successful in the solution of stiff problems have made little progress. Kong [25] performed a numerical optimization on α for fixed values of the error constant although these methods have little practical value because of their extremely small regions of accuracy about the origin. In a series of papers [40, 15, 13], Gupta and Wallace found methods with improved values for by describing LMM in terms of local polynomial approximations. The methods resulting from their investigations have larger error constants which, for many of their methods, discount

the practical value they might otherwise have. Dill and Gear [7] tried a computer-aided trial and error approach with no appreciable results. None of these investigations have resulted in methods which perform significantly better than the BDF.

We see from these reports there are several properties of LMM which appear decisive in determining the performance of a LMM in the solution of a stiff problem. To conduct a search for optimal methods we need to isolate these properties, determine a useful measure for each property, and define our object function in terms of these measures. The region of absolute stability is of prime interest because of the need to maintain stability for the components resulting from the eigenvalues with negative real part and large magnitude. The angle α is a measure of the absolute stability region which is widely used. It is a good indicator of the number of problems on which the LMM may be used. It is also a natural measure based on implementation of LMM in that stepsize is the factor which leads to difficulties in the solution of stiff problems with LMM possessing finite regions of stability. A change in stepsize moves a fixed eigenvalue radically toward or away from the origin. The angle α indicates whether such radial movement may intersect regions of instability for a given eigenvalue.

The region of accuracy about the origin is important because the components of the solution resulting from eigenvalues near the origin are dominant in the solution and their accuracy must be of concern. A natural measure for this region is the radius of relative stability. It seems clear that any measure for this region must admit

the origin as an interior point unless we preclude problems with eigenvalues of positive real part. Also, as we noted with the problem solution given in Table 4.2-2, a knowledge of the relative stability characteristics is essential before we attach meaning to the numerical solution, even for eigenvalues with negative real part. Finally the error term is a decisive property since as we see in these previous investigations, it can become prohibitively large.

These considerations led to the definition of $A(\alpha, r)$ stability given in section 2.3. It seems of interest then to know how far α and r can be extended for a fixed value of the error constant. We report the results of such an investigation in this chapter and find methods nearly optimal with regard to these desired properties. Further, in comparison we find these near-optimal methods perform successfully as we would expect from the stability and error measurements applied.

We choose the class $C(n, n)$ for our search of optimal methods. Our reasoning for this choice is similar to that used in Chapter IV when we chose the class $P(n, n)$ for our predictor search. Theorem 4.1-1 limits our investigation to classes $C(n, m)$ where $n \leq m + 1$. Other investigations, such as those previously mentioned were conducted in $C(n, n)$ and the currently popular BDF also lie in this class. These methods provide good comparisons for our results. In addition, restricting our search to $C(n, n-1)$ seems to be an unrewarding handicap. We did make several optimization runs within the class $C(4, 5)$ with no significant improvement over results we obtained from the class $C(4, 4)$. Upon consideration of these factors we conclude

C(n, n) is the better class for our search. Further, we search for A(α)-stable methods and since then A_0-stability is necessary, by virtue of theorem 2.3-7 A_∞-stability is also necessary provided we allow no roots of σ(z) to have unit magnitude. Applying the characterization of A_∞-stable methods from theorem 2.3-6, we consider only the subclass of C(n, n) for which all roots of σ(z) are of less than unit magnitude.

The corrector search is posed as an optimization problem similar to that of the predictor search in Chapter IV. The program we use is basically the same with five notable exceptions. First of all, in this problem we have two desired properties, α and r, and our object function for the optimization is defined as a linear combination of these two properties. The second difference in the two programs is sizeable and derives from the need to evaluate α. The remaining changes we refer to are those necessary for fixing the error constant, switching to correctors, and for limiting our search to A_∞-stable methods.

Defining the object function as a linear combination of α and r increases the scope of the optimizazion problem since then our interest extends to the effect of taking different linear combinations. We find an indicated relationship between the values of α and r corresponding to maximal values of these different linear combinations. The two properties are inversely related but not linearly so. For example, the following relationships are indicated to exist within C(4, 4).

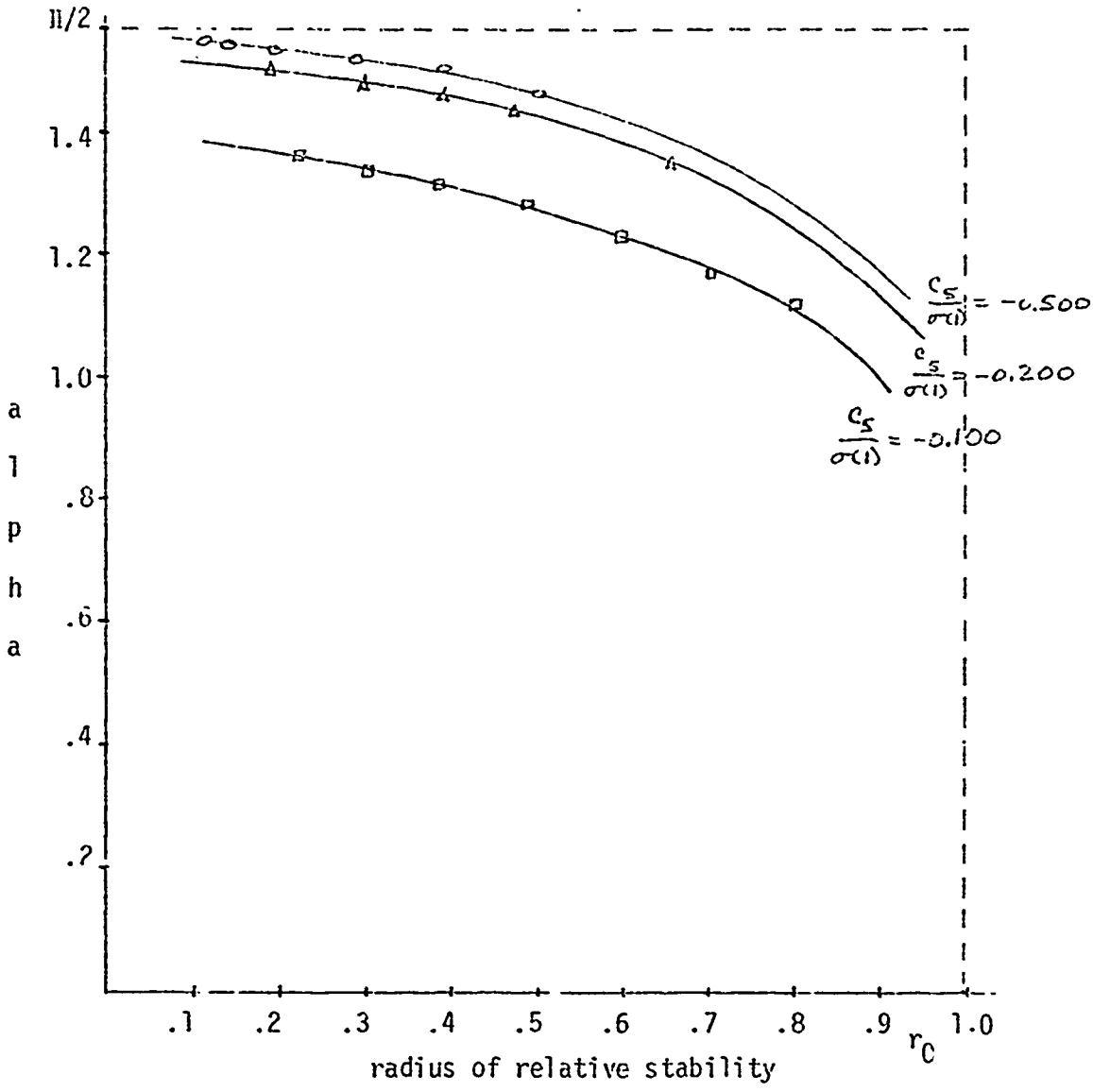


Figure 5.1-1. α vs. r.

The three curves result from fixing the error constant, $C_{p+1}/\sigma(1)$, at the values given adjacent to the curves. Notice the curves seem to approach the limiting α-value of $\pi/2$ for larger values of the radius as the error constraint is relaxed. This leads one to suspect these curves, corresponding to values of $C_{p+1}/\sigma(1)$ in $(0, \infty)$, fill in the area under the line $\alpha = \pi/2$ and to the left of

the line $r = r_0$, where r_0 is the largest radius of any method in the class C(n, n). An estimate for the value of r_0 is provided by the work of Thompson and Rodabaugh [39], where they report candidates for r_0 values for the classes C(n, n-1), $4 \leq n \leq 9$. They obtained these values by a numerical optimization with no constraint on the error constant. We did not investigate thoroughly for all orders and error constants the portion of the curve defined by linear combinations in which the radius was weighted much more than alpha. The methods in this portion of the curve are expensive to find and of little value in the solution of stiff problems because of their poor absolute stability properties. Kong [25] considered only absolute stability, and therefore his results define the upper extent of the curve on the alpha axis. The optimization runs for these methods are inexpensive, however, they also are of little value in the solution of stiff or non-stiff problems because of their poor relative stability properties. For orders greater than four we concentrate our effort in the region about and just to the left of the point where the curve starts dropping off most rapidly. This region of the curve yields methods which pair values of α and r in a way least costly to either property.

The second change is that necessary for evaluating the angle α of $A(\alpha)$ stability. We use the stability function

$$q(z) = \frac{\rho(z)}{\sigma(z)}$$

and calculate the boundary of the region of absolute stability as the image of the unit circle $z = e^{i\theta}$ at discrete points θ_n in the interval $[0, \pi]$. As reviewed in Chapter IV, for real polynomials

$p(z)$ we have $p(\bar{z}) = \overline{p(z)}$, so that $q(\bar{z}) = \overline{q(z)}$ and we need not concern ourselves with the portion of the unit circle below the real axis. Cooke [?] finds that α will be positive as long as $q(z)$ is injective and analytic in the closure of the exterior of the unit circle. We guarantee this region of analyticity by A_α-stability and the only case we find in which injectivity is violated is in crossing the real axis to the left of the origin, resulting from a method of no value in this investigation since then $\alpha = 0$. The boundary curve may have positive or negative orientation ($\text{Arg}[q(e^{2\pi i})] = \text{Arg}\,(q(1)) + 2\pi$). We impose positive orientation by assigning the calculated imaginary part of $q(z)$ to be positive. For evaluating α we are interested only in those boundary points with negative real part. From the resulting points of interest calculated, we find the one with the largest ratio,

$$\text{Imag}\,(q(e^{i\theta_n}))/\text{Re}\,(q(e^{i\theta_n})).$$

Alpha is taken as the arctangent of the negative of this largest ratio. It is important that the coefficients of ρ and σ are associated as precisely as possible to a member of $C(n, n)$. We found when poor precision was used that the resulting locus calculated was a distortion of the locus we intended to investigate. This was especially true for σ near the origin where occasionally the loci approached the origin along the real axis instead of along the imaginary axis. Within $C(n, n)$ for example, Theorem 2.2-4 guarantees this cannot happen. This distortion resulted in meaningless values for α. The use of 16-digit precision in the coefficients representing

the members of C(n, n) we investigated eliminated this problem. The routine described here to evaluate α is very fast and its cost is a relatively minor part of the optimization.

The next change we discuss is that required to restrict our investigation to the subclass of C(n, n) corresponding to a fixed error constant $C_{n+1}/\sigma(1) = E_0$. The relationship

$$\sum_{i=-1}^{n-1} i\alpha_i = \sum_{i=-1}^{n-1} \beta_i$$

derived from requiring $C_1 = 0$ in the system (1.2-3), coupled with the requirement $C_{n+1} = E_0$ yields an expression for β_{-1} explicitly in terms of $\alpha_1, \cdots, \alpha_{n-1}$ and E_0. For example if n = 3, we get

$$\beta_{-1} = [(9-24E_0) - (1 + 24E_0)\alpha_1 - 48E_0\alpha_2]/24. \qquad (5.1\text{-}1)$$

As expected, this leaves one fewer free parameters than indicated in the discussion of section 1.3. The domain of our object function for the class C(n, n) with fixed error constant is then R^{n-1}, the same as for the class P(n, n) with a floating error constant.

The last two changes referred to earlier were those required for switching to correctors and insuring A_∞-stability. Working with correctors involves inclusion of β_{-1} in calculations such as determining the coefficients of the LMM and in defining the characteristic polynomial. A_∞-stability is determined simply by checking the magnitude of the roots of σ. The optimizer is constrained away from points in the domain with roots of σ greater than one by use

of the penalty function described in [8]. The use of this constraint works very effectively in maintaining A_∞-stability.

As in section 4.1, to guarantee a positive radius of relative stability and hence zero-stability also, we consider only those methods for which $\rho(z) = (z-1)\rho_S(z)$ where $\rho_S(z)$ is a Schur polynomial. Since the coefficients of $\sigma(z)$ are determined by $\alpha_1, \cdots, \alpha_{n-1}$, and E_0, the requirement of A_∞-stability can be used to further refine the region of interest for our investigation. As an example we examine these conditions for the class C(3, 3) with fixed error constant E_0. The conditions insuring ρ_S to be Schur which follow are gained from an application of Theorem 2.1-4.

$$\begin{aligned} 1 - \alpha_1 &> 0 \\ 1 - \alpha_2 &> 0 \\ 1 + \alpha_1 + 2\alpha_2 &> 0 \end{aligned} \qquad (5.1\text{-}2)$$

Using the corollary 2.1-5 we see necessity for the polynomial $\sigma(z)$ to be Schur is similarly provided by all elements of the product vector $M_3(\beta_2, \beta_1, \beta_0, \beta_{-1})^T$ being of the same sign. Using the relationships between the α_i and β_i developed in section 1.3 and given in (5.1-1) we express the elements of the product $M_3(\beta_2, \beta_1, \beta_0, \beta_{-1})^T$ in terms of the α_i and E_0. Since two of these elements are the same as two of the expressions in (5.1-2) required positive for ρ_S to be Schur, all elements of the product vector must be positive. The resulting conditions for A_∞-stability are given in (5.1-3). The first two of these conditions supplement the conditions for zero-stability given in (5.1-2).

$$
\begin{aligned}
-(12E_0 + 1) - 12E_0\alpha_1 - (24E_0 - 1)\,\alpha_2 &> 0 \\
1 - 2\,\alpha_1 - \alpha_2 &> 0 \\
1 - \alpha_2 &> 0 \\
1 + \alpha_1 + 2\,\alpha_2 &> 0.
\end{aligned}
\qquad (5.1\text{-}3)
$$

The first inequality of (5.1-3) is the only expression involving E_0. This class of methods will always have a negative error constant since when $E_0 > 0$ the first and third inequalities in the presence of the fourth are incompatible. In general, since β_{-1} is the only parameter introducing terms involving E_0, since the coefficients of β_{-1} in each of the expressions for β_{-1}, 0, $\cdots$, β_{n-1} respectively define the first column of the matrix M_n from the proof of theorem 2.1-4, and since $M_n^2 = 2^n I_{n+1}$ we will have the error constant E_0 appearing in only one of the inequalities such as given in (5.1-3). We have not shown in general that E_0 must be negative. However, as in (5.1-1), equating the expression for C_{p+1} to $E_0 \cdot o(1)$ yields

$$
\beta_{-1} = a_0 + \sum_{i=1}^{k-1} a_i\,\alpha_i - E_0\sigma(1) \qquad (5.1\text{-}4)
$$

in general for the class C(p, k). This makes it clear that as $E_0 \to -\infty$ the inequality β_{-1} approaches the inequality $\sigma(1) > 0$, which is already included from both the zero-stability and A_∞-stability conditions. Thus as we would expect, for large negative error we gain a larger region of interest. This leads one to suspect, and from (5.1-2) and (5.1-3) it can be seen for C(3, 3), that if a given set of parameters $\alpha_1, \cdots, \alpha_{n-1}$ determine an A_∞-stable zero-stable

method for any error constant $E_0' < E_0$. However we have found counter examples to this conjecture in the general case.

5.2 Results and Comparisons

The results of the investigation described in section 1 are given here for orders four through six. The properties of the second order trapezoidal rule and the third order BDF suggest little or no improvement possible below fourth order. Since the BDF are members of C(n, n) and are used exclusively in popular stiff codes, we use them for comparison. We choose a selection of our near-optimal methods with error constants $C_{p+1}/\sigma(1)$ of -0.2 and -0.5 for fourth order, -0.4 and -0.8 for fifth order, and -0.9 for sixth order. Their stability properties are outlined in Table 5.2-1.

Table 5.2-1. Stability Properties of Near-Optimal Stiff Methods.

Order	Alpha	Radius	$C_{p+1}/\sigma(1)$
4	1.481	0.192	-0.200
	1.414	0.471	
	1.377*	0.650	
4	1.535	0.142	-0.500
	1.511	0.294	
	1.445	0.514	
5	1.431*	0.092	-0.400
	1.338	0.359	
	1.259	0.497	
5	1.485	0.077	-0.800
	1.463	0.155	
	1.394	0.463	
6	1.321*	0.121	-0.900
	1.284	0.299	
	1.065	0.435	

*Indicates use in test runs referred to later in this section.

For comparison we give corresponding information for the BDF in Table 5.2-2 below.

Table 5.2-2. Stability Properties of the BDF

Order	Alpha	Radius	$C_{p+1}/\sigma(1)$
4	1.280	0.484	-0.200
5	0.905	0.302	-0.167
6	0.311	0.130	-0.143

The k^{th} order BDF has an error constant of $1/(k+1)$. The severity of this restriction on the error constant no doubt causes the loss of A_0-stability for the higher order BDF. Notice for each order, the near-optimal methods have considerably greater regions of relative and absolute stability. The effect of the larger error constants is nullified by the capability to increase the order with no accompanying restriction on alpha. For example, if we compare the near-optimal sixth order A(1.284, 0.299)-stable method from Table 5.2-1 with the fourth order BDF, it is the case that the problem independent part of the error term for the near-optimal method will be less than that of the BDF for stepsizes less than 0.471. This particular comparison does not involve a sacrifice in the permissible values of alpha, and in other similar comparisons we could realize again in the permissible values of alpha.

We see then an increase in the error constant accompanying methods with higher values of alpha is a negative aspect that can be

controlled. However, an inferior value of alpha, such as possessed by the fifth and sixth order BDF, severely disables a method. We illustrate this fact by solving two systems of the form $y'(t) = Ay(t)$ where A is a constant 2 x 2 complex matrix and $t \in [0, 5]$. In both problems the initial values are determined by setting both constants in the general solution to 1.0. Denote the problems P1 and P2 and let them be defined respectively by the matrices A1 and A2 where

$$A1 = \begin{bmatrix} -1 & 100 \\ 0 & -100 + 373i \end{bmatrix}$$

$$A2 \quad \begin{bmatrix} -1 & 100 \\ 0 & -100 + 250i \end{bmatrix}. \qquad (5.2\text{-}1)$$

We use the constant stepsize of h = 0.005 throughout the interval [0, 5]. The calculated values are given in Appendix F. The near-optimal methods used are identified by an asterisk in Table 5.2-1. We summarize the results of these computations in Table 5.2-3 below. All calculations were done in 35-digit precision.

Table 5.2-3. Summary of Tests on Problems P1 and P2.

Order of Method	Method Used	Problem	Results
4	A(1.377, 0.650)-Stable	P1	Stable
4	BDF	P1	Unstable
5	A(1.431, 0.092)-Stable	P2	Stable
5	BDF	P2	Unstable
6	A(1.321, 0.121)-Stable	P2	Stable
6	BDF	P2	Unstable

Both P1 and B2 are complicated by the problem of stiffness. Notice the near-optimal sixth order method solved the problem P2 accurately whereas the fifth order BDF was unstable.

The difference in the relative stability radius is important as we see in the following example. Consider the problem

$$y' = -8y,\ y(0) = 1,\ x\in[0,\ 10]. \qquad (5.2\text{-}2)$$

We solve this problem with the fourth order A(1.377, 0.650)-stable method and the fourth order BDF using a constant stepsize of h = 0.1. The results are given in Table 5.2-4 which follows.

Table 5.2-4. Solution of Problem (5.2-2).

Step No.	True Solution	Fourth Order BDF		A(1.377, 0.650)-Stable	
		Calculated	Actual Error	Calculated	Actual Error
0	0.1000Q+01	0.1000Q+01	0.0	0.1000Q+01	0.0
25	0.2051Q-03	0.2070Q-06	-0.2050Q-06	-0.2179Q-08	0.4240Q-08
50	0.4248Q-17	0.3318Q-12	-0.3318Q-12	0.4131Q-17	0.1178Q-18

The previous test problems P1 and P2 demonstrated the need for methods with higher values of alpha whereas this problem demonstrates the possible effect of smaller regions of accuracy about the origin. The relative error after 50 steps in 0.02773 and $0.7811.10^5$ for the near-optimal and BDF methods, respectively.

We close with one more demonstration which accents the value of these near-optimal methods as multi-purpose methods. Lambert [28,

p. 479] indicates an interest exists in the use of stiff methods for non-stiff problems. With their large regions of accuracy about the origin these near-optimally A(α, 0)-stable methods should be ideally suited for this purpose. Consider the problem

$$y' = -y \quad y(0) = 1 \quad x \in [0, 1]. \tag{5.2-3}$$

We solve this problem with the fourth order Adams-Moulton method and the fifth order A(1.259, 0.437)-stable method from Table 5.2-1. The results given in Table 5.2-5 were obtained using a constant stepsize of 0.01.

Table 5.2-5. Solution of Problem (5.2-3)

Step No.	True Solution	4th Order Adams-Moulton Calculated	4th Order Adams-Moulton Actual Error	5th Order Near-Optimal Calculated	5th Order Near-Optimal Actual Error
0	0.10000Q+01	0.10000Q+01	0.0	0.10000Q+01	0.0
25	0.7788	0.7788	0.4557Q-10	0.7788	0.6271Q-11
50	0.6065	0.6065	0.7582Q-10	0.6065	0.1114Q-10
75	0.4724	0.4724	0.9046Q-10	0.4724	0.1356Q-10
100	0.3679	0.3679	0.9491Q-10	0.3679	0.1426Q-10

As expected, we see the large difference in error constants for these two methods (-0.02639 and -0.4 for the Adams-Moulton and near-optimal methods, respectively) is more than offset for this stepsize by using a method of one order higher. An alternative method of compensating for the larger error constant is by decreasing the stepsize but retaining the same order. We do this with the

fourth order A(1.377, 0.650)-stable method and a stepsize of 0.005. The results are given in Table 5.2-6.

Table 5.2-6. Solution of (5.2-3) by Fourth Order Near-Optimal Method.

Step	True Solution	Calculated	Actual Error
0	0.1000Q+01	0.1000Q+01	0.0
50	0.7780	0.7783	0.2269 Q-10
100	0.6065	0.6065	0.3680 Q-10
150	0.4724	0.4724	0.4356 Q-10
200	0.3679	0.3679	0.4553 Q-10

We see the error for this fourth order solution is approximately one-half that of the fourth order Adams-Moulton solution given in Table 5.2-5. From an examination of the problem independent part of the methods' error terms, we would expect more accuracy from the near-optimal method as long as the ratio of the stepsizes was less than 0.6.

In this investigation we have found a large number of near-optimal methods and it is neither practical nor necessary to list the coefficients of all these methods. For those methods listed in Table 5.2-1, we give the rational coefficients in Appendix C. In Appendix D we chart the properties of many selected methods which indicate the relationship between alpha and the radius discussed in the previous section. A relationship similar to that is indicated to exists between alpha and the radius when instead of fixing the error

constant $C_{p+1}/\sigma(1)$, we fix the error constant C_{p+1}/α_{-1}. For the class C(4, 4) we chart the latter relationships in Appendix E. It is interesting to note how well-defined these curves are, evidently indicative of the geometry of the range for the function with which we work.

There remains questions of interest in this area which have not been investigated. For example, as pertains to stiff systems, we clearly need to determine whether results as rewarding as we have found here await our investigation into higher orders. In addition, there may exist unconventional methods which could be successfully applied to large classes of stiff problems. There currently exists many computational difficulties which must be dealt with in stiff codes employing LMM. For example, since the implicit relation is solved by a form of Newton iteration, no small difficulty is presented by the need to calculate the Jacobian, especially for large systems. Even so we find LMM are the most popular methods used to solve stiff systems. This indicates great potential exists for a different approach. As pertains to non-stiff methods, there is an open question of how best to pair predictors and correctors in P-C methods. Do near-optimal predictors and near-optimal correctors pair to yield near-optimal P-C algorithms? Also as mentioned in Chapter IV, the question of how far the stability regions of predictors can be extended for fixed error constants remains open.

BIBLIOGRAPHY

1. Chase, P. E., Stability Properties of Predictor-Corrector Methods for Ordinary Differential Equations, J. Assoc. Comput. Mach., 9, 457-468 (1962).

2. Cooke, C. H., On Stiffly Stable Implicit Linear Multistep Methods, Siam J. Numer. Anal., 9, 29-34 (1972).

3. Crane, R. L., and R. W. Klopfenstein, A Predictor-Corrector Algorithm with an Increased Range of Absolute Stability, J. Assoc. Comput. Mach., 12, 227-241 (1965).

4. Cryer, C. W., A New Class of Highly Stable Methods: A0-Stable Methods, BIT, 13, 153-159 (1973).

5. Dahlquist, G., Convergence and Stability in the Numerical Integration of Ordinary Differential Equations, Math. Scand., 4, 33-53 (1956).

6. Dahlquist, G., A Special Stability Problem for Linear Multistep Methods, BIT, 3, 27-43 (1963).

7. Dill, C., and C. W. Gear, A Graphical Search for Stiffly Stable Methods for Ordinary Differential Equations, J. Assoc. Comput. Mach., 18, 75-79 (1971).

8. Fletcher, R., An Ideal Penalty Function for Constrained Optimization, J. Inst. Maths. Applics., 15, 319-342 (1975).

9. Gear, C. W., Numerical Initial Value Problems in Ordinary Differential Equations, Prentice-Hall, Englewood Cliffs, N.J., 1971.

10. Gear, C. W., The Automatic Integration of Ordinary Differential Equations, Comm. ACM., 14, 176-179 (1971).

11. Gear, C. W., Simulation: Conflicts between Real-Time and Software, In Mathematical Software III (ed. by J. R. Rice), Academic Press, New York, 1977.

12. Grigorieff, R. D., and J. Schroll, Über $A(\alpha)$-Stabile Verfahren hoher Konsistenzordnung, Computing, 20, 343-350 (1978).

13. Gupta, G. K., Some New High Order Multistep Formulae for Solving Stiff Equations, Math. Comp., 30, 417-432 (1976).

14. Gupta, G. K., A Polynomial Representation of Linear Multistep Methods, Unpublished manuscript (1979).

15. Gupta, G. K., and C. S. Wallace, Some New Multistep Methods for Solving Ordinary Differential Equations, Math. Comp., 29, 489-500 (1975).

16. Henrici, P., Discrete Variable Methods in Ordinary Differential Equations, John Wiley and Sons, 1962.

17. Henrici, P., Error Propagation for Difference Methods, John Wiley and Sons, 1963.

18. Henrici, P., Applied and Computational Complex Analysis, Vol. 1, John Wiley and Sons, 1974.

19. Isaacson, E., and H. B. Keller, Analysis of Numerical Methods, John Wiley and Sons, 1966.

20. Jeltsch, R., Stiff Stability and Its Relation to A_0- and A(0)-Stability, SIAM J. Numer. Anal., 13, 8-17 (1976).

21. Jenkins, M. A., and J. F. Traub, A Three-Stage Variable-Shift Iteration for Polynomial Zeros and Its Relation to Generalized Rayleigh Iteration, Numer. Math., 14, 252-263 (1970).

22. Jenkins, M. A., and J. F. Traub, Zeros of a Complex Polynomial, Comm. ACM, 15, 97-99 (1972).

23. Jury, E. I., Inners and Stability of Dynamic Systems, John Wiley and Sons, 1974.

24. Klopfenstein, R. W., and R. S. Millman, Numerical Stability of a One-Evaluation Predictor-Corrector Algorithm for Numerical Solution of Ordinary Differential Equations, Math. Comp. 22, 557-564 (1968).

25. Kong, A. K., A Search for Better Linear Multistep Methods for Stiff Problems, Dissertation, University of Illinois at Urbana-Champaign, 1977.

26. Krall, A. M., Stability Techniques for Continuous Linear Systems, Gordon and Breach Science Publishers, New York, 1967.

27. Lambert, J. D., Computational Methods in Ordinary Differential Equations, John Wiley and Sons, 1973.

28. Lambert, J. D., The Initial Value Problem for Ordinary Differential Equations, In The State of the Art in Numerical Analysis (ed. by D. Jacobs), Academic Press, 1977.

29. Marden, M., Geometry of Polynomials, American Mathematical Society, Providence, Rhode Island, 1966.

30. Odeh, F., and W. Liniger, A Note on Unconditional Fixed-h Stability of Linear Multistep Formulae, Computing, 7, 240-253 (1971).

31. Powell, M. J. D., An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives, Comput. J., 7, 155-162 (1964).

32. Powell, M. J. D., A Method for Nonlinear Constraints in Minimization Problems, In Optimization (ed. by R. Fletcher), Academic Press, 1969.

33. Ralston, A., A First Course in Numerical Analysis, McGraw-Hill, 1965.

34. Rodabaugh, D. J., On Stable Correctors, Comput. J., 13, 98-100 (1970).

35. Rodabaugh, D. J., and S. Thompson, Optimally Stable Fifth Order Integration Methods: A Numerical Approach, Int. J. Math. and Math. Sci., 1, 319-334 (1978).

36. Rodabaugh, D. J., and J. R. Wesson, On the Efficient Use of Predictor-Corrector Methods in the Numerical Solution of Differential Equations, NASA TN D-2946, 1-40 (1965).

37. Shampine, L. F., Implementation of Implicit Formulas for the Solution of ODEs, SAND 79-0694, Sandia Laboratories, New Mexico, 1979.

38. Skeel, R. D., Equivalent Forms of Multistep Formulas, Dept. of Computer Science Report, University of Illinois, 1978.

39. Thompson, S., and D. J. Rodabaugh, Corrector Methods with Increased Ranges of Stability, Comput. and Math. with App., 3, 197-201 (1977).

40. Wallace, C. S., and G. K. Gupta, General Linear Multistep Methods to Solve Ordinary Differential Equations, Austral. Comput. J., 5, 62-69 (1973).

41. Widlund, O. B., A Note on Unconditionally Stable Linear Multistep Methods, BIT, 7, 65-70 (1967).

42. Willoughby, R. A., editor, Stiff Differential Systems, Plenum Press, 1974.

APPENDIX A

COEFFICIENT MATRICES

In the matrices given below we list only the numerator of the entries. All entries of the matrix for C(n, n) have the common denominator D_n.

C(2, 2), $D_2 = 2$

$$\begin{bmatrix} 3 & 1 & -4 \\ -1 & 1 & 2 \end{bmatrix}$$

C(3, 3) $D_3 = 12$

$$\begin{bmatrix} 23 & 5 & 4 & -36 \\ -16 & 8 & 16 & 36 \\ 5 & -1 & 4 & -12 \end{bmatrix}$$

C(4, 4) $D_4 = 24$

$$\begin{bmatrix} 55 & 9 & 8 & 9 & -96 \\ -59 & 19 & 32 & 27 & 144 \\ 37 & -5 & 8 & 27 & -96 \\ -9 & 1 & 0 & 9 & 24 \end{bmatrix}$$

C(5, 5), $D_5 = 720$

$$\begin{bmatrix} 1901 & 251 & 232 & 243 & 224 & -3600 \\ -2774 & 646 & 992 & 918 & 1024 & 7200 \\ 2616 & -264 & 192 & 648 & 384 & -7200 \\ -1274 & 106 & 32 & 378 & 1024 & 3600 \\ 251 & -19 & -8 & -27 & 224 & -720 \end{bmatrix}$$

C(6, 6), D_6 = 1440

$$\begin{bmatrix} 4277 & 475 & 448 & 459 & 448 & 475 & -8640 \\ -7923 & 1427 & 2064 & 1971 & 2048 & 1875 & 21600 \\ 9982 & -798 & 224 & 1026 & 768 & 1250 & -28900 \\ -7298 & 482 & 224 & 1026 & 2048 & 1250 & 21600 \\ 2877 & -173 & -96 & -189 & 448 & 1875 & 8640 \\ -475 & 27 & 16 & 27 & 0 & 475 & 1440 \end{bmatrix}$$

C(7, 7), D_7 = 60480

$$\begin{bmatrix} 198721 & 19087 & 18224 & 18495 & 18304 & 18575 & 17712 & -423360 \\ -447288 & 65112 & 90240 & 87480 & 89088 & 87000 & 93312 & 1270080 \\ 705549 & -46461 & 528 & 31347 & 24576 & 31875 & 11664 & -2116800 \\ -688256 & 37504 & 21248 & 58752 & 96256 & 80000 & 117504 & 2116800 \\ 407139 & -20211 & -12912 & -19683 & 11136 & 58125 & 11664 & -1270080 \\ -134472 & 6312 & 4224 & 5832 & 3072 & 28200 & 93312 & 423360 \\ 19087 & -863 & -592 & -783 & -512 & -1375 & 17712 & -60480 \end{bmatrix}$$

C(8, 8), D_8 = 120960

434241	36799	35424	35775	35584
-1152169	139849	187648	183465	185344
2183877	-121797	-20448	37179	27648
-2664477	123133	78336	160029	228352
2102243	-88547	-61664	-81891	-13568
-1041723	41499	29952	37179	27648
295767	-11351	8352	-10071	-8192
-36799	1375	1024	1215	1024

35775	35424	36799	-967680
183625	186624	175273	3386880
34875	23328	64827	-6773760
208125	235008	146461	8457200
68125	23328	146461	-6773760
85275	186624	64827	3386880
-12735	35424	175273	-967680
1375	0	36799	120960

C(9, 9), D_9 = 3628800

14 097 247	1 070 017	1 036 064	1 043 361	1 040 128
-43 125 206	4 467 094	5 842 688	5 743 062	5 779 456
95 476 786	-4 604 594	-1 359 808	278 478	62 464
-139 855 262	5 595 358	3 842 816	6 474 654	8 384 512
137 968 480	-5 033 120	-3 715 840	-4 548 960	-2 324 480
-91 172 642	3 146 338	2 391 296	2 789 154	2 363 392
38 833 486	-1 291 214	- 996 928	-1 139 022	-1 012 736
-9 664 106	312 874	243 968	275 562	249 856
1 070 017	-33 953	-26 656	-29 889	-27 392

1 042 625	1 039 392	1 046 689	1 012 736	-32 659 200
5 753 750	5 785 344	5 716 438	6 029 312	130 636 800
188 750	46 656	340 942	- 950 272	-340 819 200
7 958 750	8 356 608	7 601 566	10 747 904	457 228 800
- 100 000	- 933 120	384 160	-4 648 960	-457 228 800
4 273 250	6 905 088	5 152 546	10 747 904	304 819 200
-1 228 750	409 536	3 654 322	- 950 272	-130 636 800
286 250	186 624	1 562 218	6 029 312	32 659 200
-30 625	-23 328	-57 281	1 012 736	-3 628 800

APPENDIX B

PREDICTOR COEFFICIENTS

The following coefficients are given for the methods referred to in Table 4.2-1. The numerators of the coefficients are listed in the tables below and each order has a common denominator denoted by D.

Table B-1.

Coefficient	4th Order	5th Order	6th Order
α_{-1}	-300	-720 000 000	-14 400 000
α_0	591	1 705 707 360	40 006 080
α_1	-462	-1 697 984 640	-39 241 440
α_2	201	987 585 120	17 628 480
α_3	-30	-297 991 440	-8 373 600
α_4		22 683 600	6 521 760
α_5			-2 141 280
β_0	570	1 533 770 569	33 963 773
β_1	-869	-3 284 474 686	-97 823 787
β_2	592	3 245 856 024	119 961 838
β_3	-143	-1 604 272 786	-81 922 322
β_4		303 066 559	32 649 093
β_5			-6 153 235
D	300	720 000 000	14 400 000

Table B-2.

Coefficient	7th Order	8th Order	9th Order
α_{-1}	-604 800 000	-1 209 600 000	-36 288 000 000
α_0	1 479 522 240	2 352 430 080	68 929 056 000
α_1	-1 562 984 640	-2 067 206 400	-57 556 396 800
α_2	1 112 045 760	1 677 473 280	44 478 201 600
α_3	-599 477 760	-895 950 720	-29 756 160 000
α_4	252 383 040	456 624 000	25 358 054 400
α_5	-76 688 640	-411 022 080	-13 684 204 800
α_6	0	12 096 000	-2 474 841 600
α_7		85 155 840	232 243 200
α_8			762 048 000
β_0	1 698 536 399	3 982 004 743	131 051 879 512
β_1	-5 101 980 072	-12 450 561 759	-461 208 069 296
β_2	8 017 316 643	23 395 159 971	1 008 365 805 136
β_3	-7 743 202 432	-28 566 654 731	-1 467 668 483 312
β_4	4 524 155 853	22 109 841 909	1 435 089 863 680
β_5	-1 510 919 256	-11 107 551 069	-956 912 924 432
β_6	209 655 425	3 248 488 993	403 247 072 176
β_7		-361 187 577	-98 967 442 256
β_8			11 306 055 592
D	604 800 000	1 209 600 000	36 288 000 000

APPENDIX C

CORRECTOR COEFFICIENTS

The following coefficients are given for the methods referred to in Table 5.2-1. The numerators of the coefficients are listed in the table below. The coefficients of each method have a common denominator denoted by D.

Table C-1. Fourth Order, $C_5/o(1) = -0.200$.

Coefficients	A(1.481, 0192)-Stable	A(1.414, 0.471)-Stable	A(1.377, 0.650)-Stable
α_{-1}	-30 000 000	-900	-240 000
α_0	60 267 000	1800	464 400
α_1	-38 319 000	-1260	-306 000
α_2	1 743 000	360	87 288
α_3	6 309 000	0	-5 088
β_{-1}	14 032 375	415	109 512
β_0	1 197 750	50	24 165
β_1	-10 470 000	-240	-64 993
β_2	5 782 250	110	19 199
β_3	3 551 625	25	4 829
D	30 000 000	900	240 000

Table C-2. Fourth Order, $C_5/o(1) = -0.500$.

Coefficients	A(1.535, 0.142)-Stable	A(1.511, 0.294)-Stable	A(1.445, 0.514)-Stable
α_{-1}	-900 000 000	-900 000	-90 000 000
α_0	2 312 370 000	2 448 000	202 104 000
α_1	-2 057 940 000	-2 520 000	-160 425 000
α_2	630 900 000	1 170 000	53 982 000
α_3	14 670 000	-198 000	-5 661 000
β_{-1}	434 431 625	437 675	45 498 925
β_0	-231 152 750	-317 450	-20 035 950
β_1	-377 409 000	-244 200	-9 652 200
β_2	305 314 750	329 050	1 801 550
β_3	16 685 375	-79 075	2 941 675
D	900 000 000	900 000	90 000 000

Table C-3 Fifth Order, $C_6/\sigma(1) = -0.400$.

Coefficients	A(1.431, 0.092)-Stable	A(1.338, 0.359)-Stable	A(1.259, 0.514)-Stable
α_{-1}	-9 000 000	-9 000 000	-720 000
α_0	27 720 000	24 120 000	1 809 000
α_1	-35 100 000	-26 100 000	-1 728 000
α_2	22 500 000	14 400 000	792 000
α_3	-7 200 000	-3 600 000	-144 000
α_4	1 080 000	180 000	0
β_{-1}	4 159 875	4 013 625	321 408
β_0	-4 117 125	-1 923 375	-101 840
β_1	-1 212 750	-2 450 250	-202 720
β_2	4 067 250	2 829 750	117 120
β_3	-1 537 125	-693 375	38 240
β_4	259 875	-156 375	-28 208
D	9 000 000	9 000 000	720 000

Table C-4. Fifth Order, $C_0/\sigma(1) = -0.800$.

Coefficients	A(1.485, 0.077)-Stable	A(1.463, 0.155)-Stable	A(1.394, 0.463)-Stable
α_{-1}	-9 000 000	-14 400	-9 000 000
α_0	28 800 000	44 640	27 352 800
α_1	-36 000 000	-54 720	-33 753 600
α_2	20 700 000	31 680	21 636 900
α_3	-4 500 000	-7 200	-7 236 000
α_4	0	0	999 900
β_{-1}	4 218 125	6 711	4 146 335
β_0	-4 726 875	-6 833	-3 895 235
β_1	-2 011 250	-2 998	-1 499 110
β_2	5 188 750	7 242	3 40. 690
β_3	-1 576 875	-2 353	-1 577 785
β_4	-191 875	-329	223 905
D	9 000 000	14 400	9 000 000

Table C-5. Sixth Order, $C_7/\sigma(1) = -0.900$.

Coefficient	A(1.321, 0.121)-Stable	A(1.284, 0.299)-Stable	A(1.065, 0.435)-Stable
α_{-1}	-14 400 000	-604 800 000	-604 800
α_0	49 824 000	2 047 066 500	1 959 552
α_1	-72 000 000	-2 903 040 000	-2 721 600
α_2	53 280 000	2 177 280 000	2 116 800
α_3	-18 720 000	-846 720 000	-967 680
α_4	1 440 000	117 996 480	241 920
α_5	576 000	12 216 960	-24 192
β_{-1}	6 563 520	264 974 986	274 447
β_0	-9 114 120	-302 886 000	-330 695
β_1	1 415 800	-84 322 494	114 147
β_2	4 667 600	310 644 705	-40 448
β_3	-1 129 200	-148 424 994	103 437
β_4	-1 858 120	-14 144 400	-81 720
β_5	894 520	23 388 886	21 313
D	14 400 000	604 800 000	604 800

APPENDIX D

RELATIONSHIP OF ALPHA VS. RADII FOR FIXED GLOBAL TRUNCATION ERROR CONSTANTS

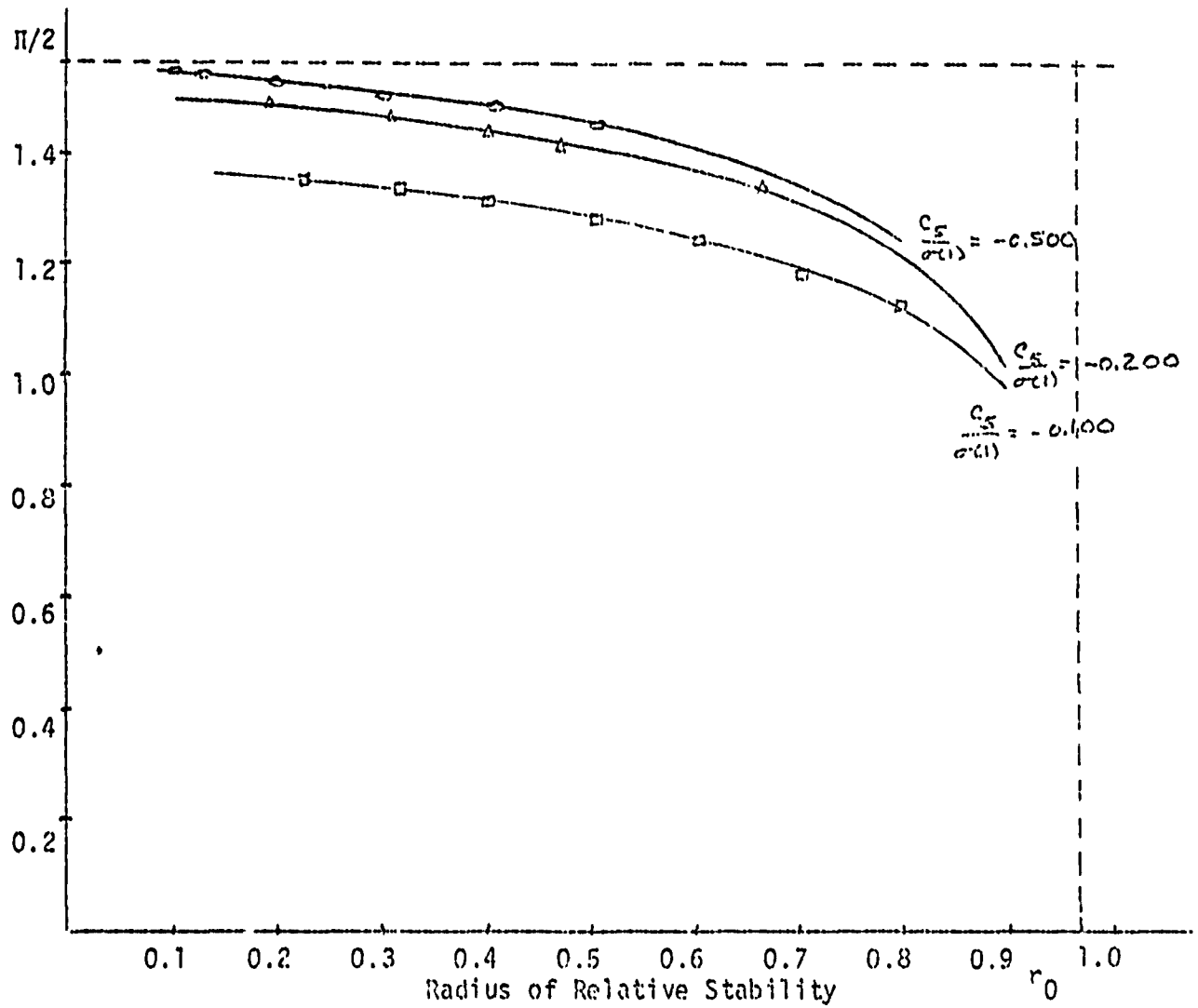


Figure 0-1. Relationship for C(4, 4).

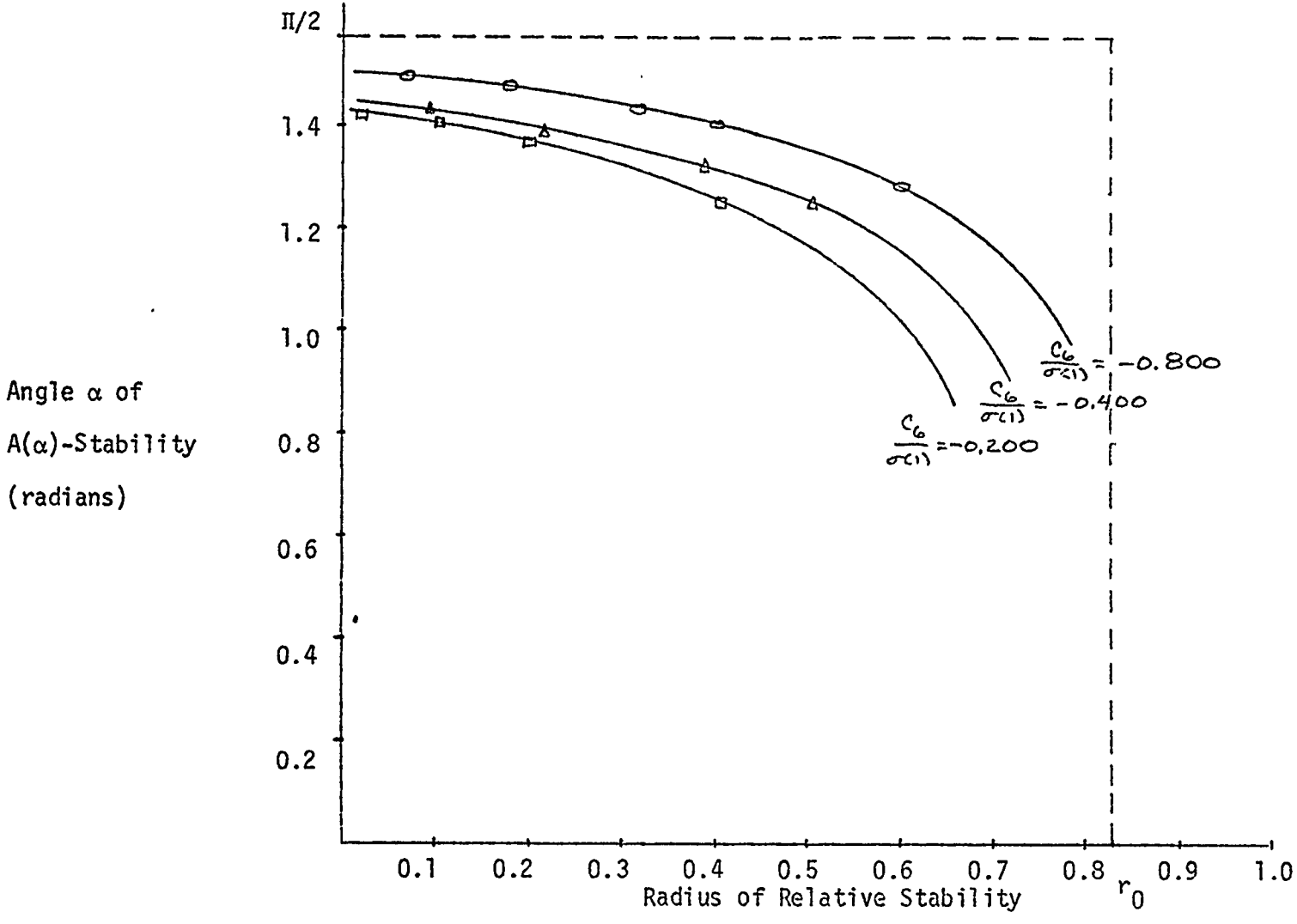


Figure D-2. Relationship for C(5, 5).

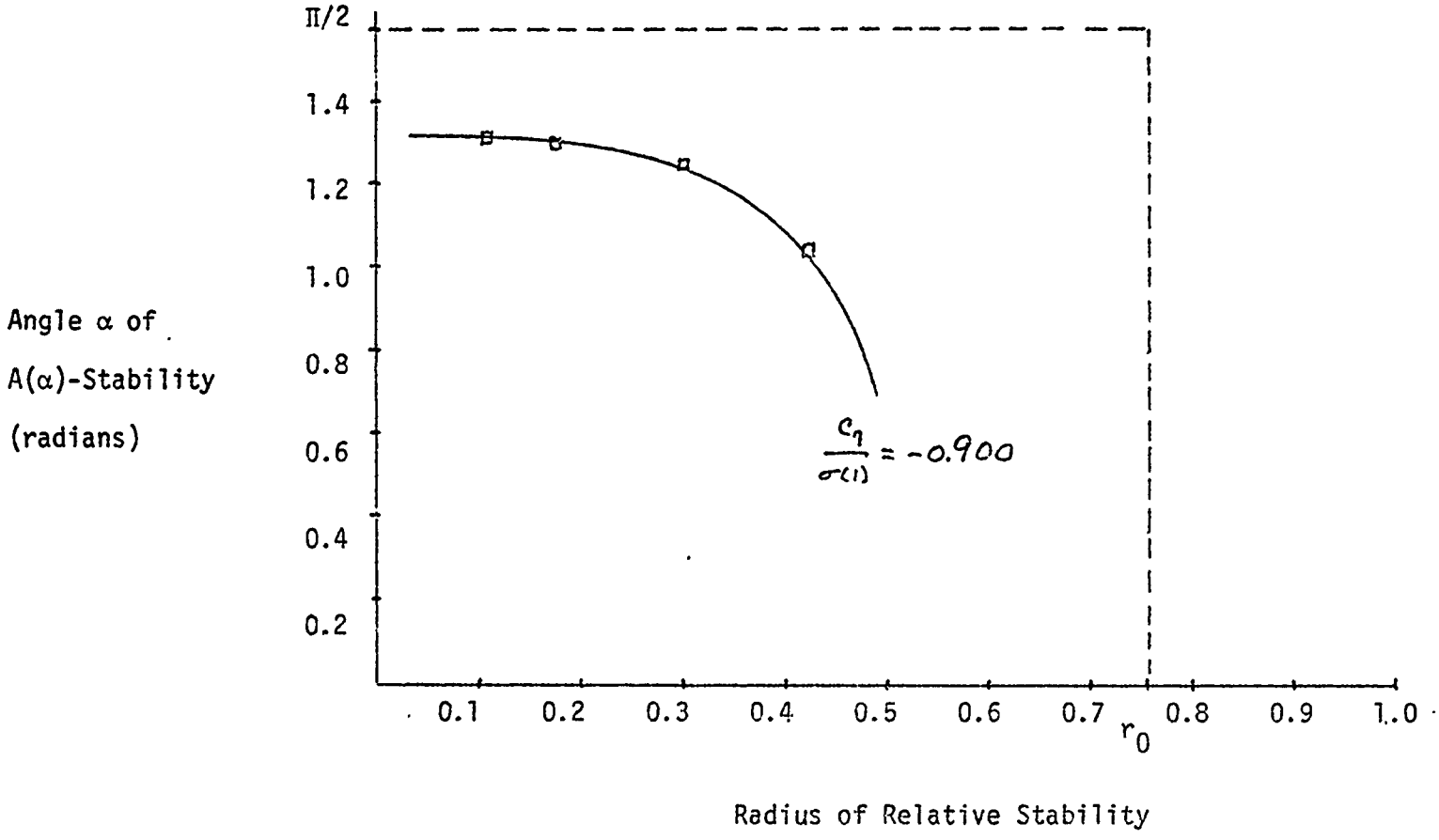


Figure D-3. Relationship for C(6, 6).

APPENDIX E

RELATIONSHIP OF ALPHA VS. RADIUS FOR FIXED LOCAL TRUNCATION ERROR CONSTANTS

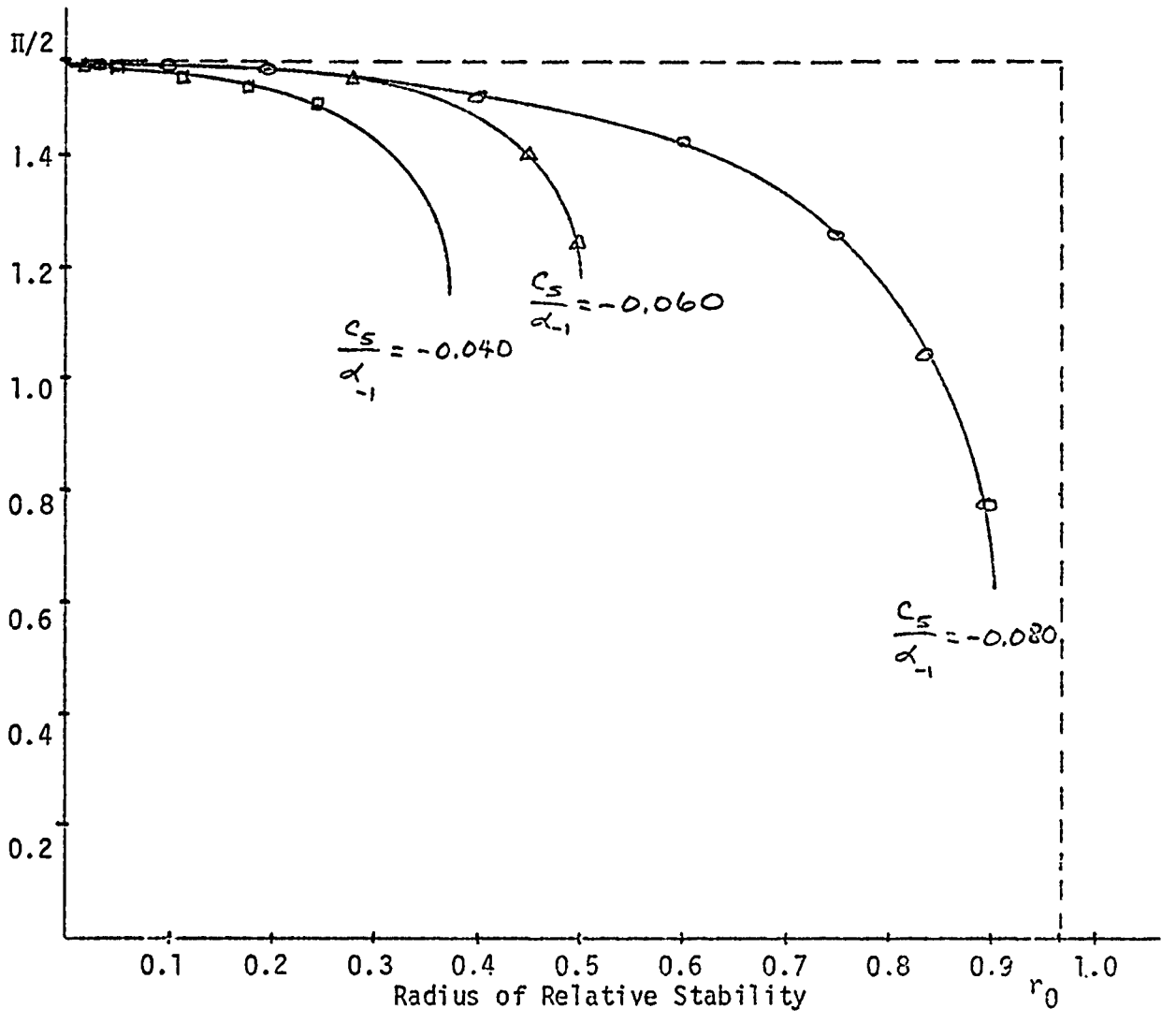


Figure E-1. Relationship for C(4, 4).

APPENDIX F

SOLUTION OF TEST PROBLEMS P1 AND P2

Calculations below were carried out on an Amdahl 370/V7 in 35-digit precision. Numbers less than approximately 10^{-78} in magnitude are represented as 0.0.

Table F-1. True Solution of Problem P1.

Step (h=0.005)	First Component		Second Component	
	Real	Imaginary	Real	Imaginary
0	0.2000Q+01	0.0	-0.9900Q+00	0.3730Q+01
200	0.3679Q+00	0.2794Q-43	-0.7988Q-43	-0.1193Q-42
400	0.1353Q+00	0.0	0.0	0.0
600	0.4979Q-01	0.0	0.0	0.0
800	0.1832Q-01	0.0	0.0	0.0
1000	0.6738Q-02	0.0	0.0	0.0

Table F-2. Calculated Solution of P1 by Fourth Order A(1.377, 0.650)--Stable Method.

Step (h=0.005)	First Component		Second Component	
	Real	Imaginary	Real	Imaginary
0	0.2000Q+01	0.0	-0.9900Q+00	0.3730Q+01
200	0.3679Q+00	-0.1916Q-05	0.6626Q-05	0.3862Q-05
400	0.1353Q+00	0.4772Q-11	-0.1198Q-10	-0.2664Q-10
600	0.4979Q-01	-0.1773Q-17	-0.2188Q-16	0.1091Q-15
800	0.1832Q-01	-0.5826Q-22	0.3095Q-21	-0.2697Q-21
1000	0.6738Q-02	0.3869Q-27	-0.1601Q-26	0.2124Q-27

Table F-3. Actual Global Error of Calculated Solution of P1 by Fourth Order A(1.377, 0.650)-Stable Method.

Step	First Component		Second Component	
	Real	Imaginary	Real	Imaginary
0	0.0	0.0	0.0	0.0
200	-0.5267Q-06	0.1916Q-05	-0.6626Q-05	-0.3862Q-05
400	0.3970Q-10	-0.4772Q-11	0.1198Q-10	0.2664Q-10
600	0.1872Q-10	0.1773Q-17	0.2118Q-16	-0.1019Q-15
800	0.9199Q-11	0.5826Q-22	-0.3095Q-21	0.2897Q-21
1000	0.4234Q-11	-0.3869Q-27	0.1601Q-26	-0.2124Q-27

Table F-4. Calculated Solution of P1 by Fourth Order BDF.

Step (h=0.005)	First Component		Second Component	
	Real	Imaginary	Real	Imaginary
0	0.2000Q+01	0.0	-0.9900Q+00	0.3730Q+01
200	0.9475Q+01	0.5968Q+01	-0.3128Q+02	0.2806Q+02
400	0.2122Q+03	-0.1348Q+03	0.2928Q+03	0.9245Q+03
600	-0.8181Q+03	-0.5742Q+04	0.2223Q+05	-0.2633Q+04
800	-0.1285Q+06	-0.3738Q+05	0.2669Q+06	-0.4424Q+06
1000	-0.2008Q+07	0.2348Q+07	-0.6769Q+07	-0.9815Q+07

Table F-5. Actual Global Error of Calculated Solution of P1 by Fourth Order BDF.

Step	First Component		Second Component	
	Real	Imaginary	Real	Imaginary
0	0.0	0.0	0.0	0.0
200	-0.9107Q+01	-0.5968Q+01	0.3128Q+02	-0.2806Q+02
400	-0.2121Q+03	0.1348Q+03	-0.2128Q+03	-0.9245Q+03
600	0.8182Q+03	0.5742Q+04	-0.2223Q+05	-0.2533Q+04
800	0.1285Q+06	0.3738Q+05	-0.2669Q+06	0.4424Q+06
1000	0.2008Q+07	-0.2348Q+07	0.6769Q+07	0.9815Q+07

Table F-6. True Solution of Problem P2.

Step (h=0.005)	First Component		Second Component	
	Real	Imaginary	Real	Imaginary
0	0.200Q+01	0.0	-0.9900Q+00	0.2500Q+01
200	0.3679Q+00	-0.3610Q-43	0.8139Q-43	0.5816Q-43
400	0.1353Q+00	0.0	0.0	0.0
600	0.4979Q-01	0.0	0.0	0.0
800	0.1832Q-01	0.0	0.0	0.0
1000	0.6738Q-02	0.0	0.0	0.0

Table F-7. Calculated Solution of P2 by Fifth Order A(1.431, 0.092)-Stable Method.

Step (h=0.005)	First Component		Second Component	
	Real	Imaginary	Real	Imaginary
0	0.2000Q+01	0.0	-0.9900Q+00	0.2500Q+01
200	0.3679Q+00	-0.5067Q-12	-0.8977Q-12	0.5967Q-11
400	0.1353Q+00	-0.4956Q-23	0.7781Q-22	-0.1604Q-21
600	0.4979Q-01	-0.1757Q-31	-0.4609Q-32	0.4982Q-32
800	0.1832Q-01	-0.6335Q-32	0.2243Q-42	-0.1293Q-42
1000	0.6738Q-02	-0.1333Q-32	-0.9632Q-53	0.2167Q-53

Table F-8. Actual Global Error of Calculated Solution of P2 by Fifth Order A(1.431, 0.092)-Stable Method.

Step	First Component		Second Component	
	Real	Imaginary	Real	Imaginary
0	0.0	0.0	0.0	0.0
200	-0.2413Q-11	0.5067Q-12	0.8977Q-12	-0.5967Q-11
400	-0.1687Q-12	0.4946Q-23	-0.7781Q-22	0.1604Q-21
600	-0.9338Q-13	0.1757Q-31	0.4609Q-32	-0.4982Q-32
800	-0.4588Q-13	0.6335Q-32	-0.2243Q-42	0.1293Q-42
1000	-0.2112Q-13	0.1333Q-32	0.9632Q-53	-0.2167Q-53

Table F-9. Calculated Solution of P2 by Fifth Order BDF.

Step (h=0.005)	First Component		Second Component	
	Real	Imaginary	Real	Imaginary
0	0.2000Q+01	0.0	-0.9900Q+00	0.2500Q+01
200	0.9365Q+03	0.2558Q+04	-0.7321Q+04	-0.1914Q+03
400	0.1696Q+08	0.6429Q+08	-0.1775Q+09	-0.2124Q+08
600	0.2666Q+12	0.1601Q+13	-0.4267Q+13	-0.9187Q+12
800	0.2852Q+16	0.3953Q+17	-0.1016Q+18	-0.3200Q+17
1000	-0.2026Q+20	0.9673Q+21	-0.2398Q+22	-0.1008Q+22

Table F-10. Actual Global Error of Calculated Solution of P2 by Fifth Order BDF.

Step	First Component		Second Component	
	Real	Imaginary	Real	Imaginary
0	0.0	0.0	0.0	0.0
200	-0.9362Q+03	-0.2558Q+04	0.7321Q+04	0.1914Q+03
400	-0.1696Q+08	-0.6429Q+08	0.1775Q+09	0.2124Q+08
600	-0.2666Q+12	-0.1601Q+13	0.4267Q+13	0.9187Q+12
800	-0.2852Q+16	-0.3953Q+17	0.1016Q+18	0.3200Q+17
1000	0.2026Q+20	-0.9673Q+21	0.2398Q+22	0.1008Q+22

Table F-11. Calculated Solution of P2 by Sixth Order A(1.321, 0.121)-Stable Method.

Step (h=0.005)	First Component		Second Component	
	Real	Imaginary	Real	Imaginary
0	0.2000Q+01	0.0	-0.9900Q+00	0.2500Q+01
200	0.3679Q+00	0.6007Q-10	-0.4715Q-09	0.7519Q-09
400	0.1353Q+00	-0.2089Q-18	0.4503Q-18	0.3885Q-18
600	0.4979Q-01	-0.6950Q-28	0.3034Q-27	-0.2590Q-27
800	0.1832Q-01	-0.2182Q-31	-0.1408Q-36	-0.2275Q-36
1000	0.6738Q-02	-0.5375Q-32	-0.1651Q-45	0.7022Q-46

Table F-12. Actual Global Error of Calculated Solution of P2 by Sixth Order A(1.321, 0.121)-Stable Method.

Step	First Component		Second Component	
	Real	Imaginary	Real	Imaginary
0	0.0	0.0	0.0	0.0
200	-0.3246Q-09	-0.6007Q-10	0.4715Q-09	-0.7519Q-09
400	0.3810Q-14	0.2089Q-18	-0.4503Q-18	-0.3885Q-18
600	0.2113Q-14	0.6950Q-28	-0.3034Q-27	0.2590Q-27
800	0.1039Q-14	0.2182Q-31	0.1408Q-36	0.2275Q-36
1000	0.4786Q-15	0.5375Q-32	0.1651Q-45	-0.7022Q-46

Table F-13. Calculated Solution of P2 by Sixth Order BDF.

Step (h=0.005)	First Component		Second Component	
	Real	Imaginary	Real	Imaginary
0	0.2000Q+01	0.0	-0.9900Q+00	0.2500Q+01
200	-0.1667Q+14	-0.4319Q+14	0.1245Q+15	0.1089Q+13
400	0.3979Q+29	0.5945Q+29	-0.1880Q+30	0.4061Q+29
600	-0.8016Q+44	-0.7611Q+44	0.2696Q+45	-0.1250Q+45
800	0.1467Q+60	0.8753Q+59	-0.3640Q+60	0.2800Q+60
1000	-0.2508Q+75	-0.8216Q+74	0.4537Q+75	-0.5457Q+75

Table F-14. Actual Global Error of Calculated Solution of P2 by Sixth Order BDF.

Step	First Component		Second Component	
	Real	Imaginary	Real	Imaginary
0	0.0	0.0	0.0	0.0
200	0.1667Q+14	0.4319Q+14	-0.1245Q+15	-0.1089Q+13
400	-0.3979Q+29	-0.5945Q+29	0.1880Q+30	-0.4061Q+29
600	0.8016Q+44	0.7611Q+44	-0.2696Q+45	0.1250Q+45
800	-0.1467Q+60	-0.8753Q+59	0.3640Q+60	-0.2800Q+60
1000	0.2508Q+75	0.8216Q+74	-0.4537Q+75	0.5457Q+75